

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 02:39:19 ; Search time 3112 Seconds
(without alignments)
7191.535 Million cell updates/sec

Title: US-09-555-342B-1_COPY_1733_2501
Perfect score: 769
Sequence: 1 tcaactcgtggttcagagc.....attgaggagagcgaagca 769

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl :

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vt.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vt.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_mus.*
- 34: em_hgt_pln.*
- 35: em_hgt_rod.*
- 36: em_hgt_mam.*
- 37: em_hgt_vrt.*
- 38: em_sy.*
- 39: em_hgt_hum.*
- 40: em_hgt_mus.*
- 41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	769	100.0	3442	9	AB008430	AB008430 Homo sapi
2	631.4	82.1	3902	10	BC030329	BC030329 Mus muscu
3	478.8	62.3	2632	10	BC004009	BC004009 Mus muscu
4	312.8	40.7	2431	10	BC027077	BC027077 Mus muscu
5	280.8	36.5	3997	9	AB018336	AB018336 Homo sapi
6	271.2	35.3	3719	10	BC009153	BC009153 Mus muscu
7	241	31.3	106578	9	AL137249	AL137249 Human DNA
8	204.2	26.6	224310	2	AC094777	AC094777 Rattus no
9	199.4	25.9	221502	2	AC122885	AC122885 Mus muscu
10	162.4	21.1	393	11	G60059	G60059 SHGC-130885
11	136.8	17.8	152053	2	AC022669	AC022669 Homo sapi
12	120.8	15.7	426	6	AX261839	AX261839 Sequence
13	119.2	15.5	190588	2	AC101725	AC101725 Mus muscu
14	119.2	15.5	221502	2	AC122885	AC122885 Mus muscu
15	111.8	14.5	149417	2	AC118779	AC118779 Rattus no
16	97.2	12.6	42014	2	AC018327	AC018327 Drosophil
17	97.2	12.6	160710	3	AC009537	AC009537 Drosophil
18	97.2	12.6	164713	3	AC007589	AC007589 Drosophil
19	97.2	12.6	303209	3	AE003604	AE003604 Drosophil
20	97	12.6	62537	2	AC101100	AC101100 Mus muscu
21	95.8	12.5	105605	2	AC115336	AC115336 Rattus no
22	95.8	12.5	176630	2	AC109427	AC109427 Rattus no
23	95	12.4	3055	10	BC026860	BC026860 Mus muscu
24	87.8	11.4	134911	9	AC005104	AC005104 Homo sapi
25	84.2	10.9	152053	2	AC022669	AC022669 Homo sapi
26	83.6	10.9	1286	6	AX411309	AX411309 Sequence
27	83.6	10.9	1821	9	AK002045	AK002045 Homo sapi
28	83.6	10.9	3610	6	AX405929	AX405929 Sequence
29	83.6	10.9	3842	9	AB037783	AB037783 Homo sapi
30	82	10.7	1478	9	BC013319	BC013319 Homo sapi
31	81.4	10.6	3220	5	AF017370	AF017370 Danio rer
32	80.4	10.5	120911	2	AC130919	AC130919 Rattus no
33	80.2	10.4	67896	2	AC131316	AC131316 Mus muscu
34	79.6	10.4	133745	2	AC131454	AC131454 Strongylo
35	73.8	9.6	2605	9	BC021301	BC021301 Homo sapi
36	72.8	9.5	462	6	AX371287	AX371287 Sequence
37	70.8	9.2	138025	9	HSD120N4	AL096709 Human DNA
38	65.8	8.6	2956	10	AF017369	AF017369 Mus muscu
39	64.4	8.4	1896	9	BC032232	BC032232 Homo sapi
40	64.4	8.4	3224	9	BC032429	BC032429 Homo sapi
41	64.4	8.4	4670	9	AK000004	AK000004 Homo sapi
42	62.2	8.1	2170	9	AK097217	AK097217 Homo sapi
43	62.2	8.1	120911	2	AC130919	AC130919 Rattus no
44	62.2	8.1	149417	2	AC118779	AC118779 Rattus no
45	62.2	8.1	224310	2	AC094777	AC094777 Rattus no

ALIGNMENTS

RESULT 1	AB008430	Homo sapiens mRNA for CDEP, complete cds.	3442 bp	mRNA	linear	PRI 13-FEB-1999
LOCUS	AB008430	Homo sapiens				
DEFINITION	AB008430	Homo sapiens				
ACCESSION	AB008430	Homo sapiens				
VERSION	AB008430.1	GI:2766164				
KEYWORDS	CDEP.					
SOURCE	Homo sapiens	embryo cartilage chondrocyte cDNA to mRNA.				
ORGANISM	Homo sapiens					
REFERENCE	1 (sites)	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS	Koyano,Y., Kawamoto,T., Shen,M., Yan,W., Noshiro,M., Fujii,K. and Kato,Y.	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE		Molecular cloning and characterization of CDEP, a novel human				

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 65 Row: m Column: 19
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

Location/Qualifiers
 1. .3902
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N-3"
 /clone="IMAGE:5376197"
 /tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."
 /clone_lib="NCI_CGAP_Mam2"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 <1..2510
 /codon_start=3
 /product="Unknown (protein for IMAGE:5376197)"
 /protein_id="AAH30329.1"
 /db_xref="GI:20987936"
 /translation="TRREVARRLMYGIRLHPAKDRGKTKINLAVANTGILVFGPTK
 INANWAKRKLSPKRRFLIKRPDWNSSQYDLEFLMAGRDECFKFWKICVHHAP
 FRLEEPKPKPVLFRGSGFRSGRTQKQVLDYVREGHKYQFERKSKIHSTRS
 LVSPQAPNSVPPQSSSLTFEGTESPGQSCQOAKETKACTLELGHQSPALP
 KSPGSKAADGTTVPPEEEEREGKDGIRPSNPPOPSTGSLTSGPHSELINS
 OGGAAPANVTLNPLSPDNKOASPLISPLNDQACPTDDEEERRRFPDIXAYIA
 KEVSTERTYIKLDELVIASFPQSVREDSMPKLSLIFNPEFLPHFHNFKUEIE
 QRLALWGRNSHVRGDIYRIGDYMKNQIMKLAHLKWLKSHSEALAEFSIKGRR
 LEHCRDFEQLKQVCLPLNTFLRLHMLHYKHLVLEKLPKHPNHNADFCRAALA
 EITEMVAQLHGTMIKMFNFQKLHKKLDLIGIDNLVTPGREFIRLGSLSLKGGLQ
 RMFFLNDVLLYTRGLTASNOFVGOLPLYGMTIESEBEWGVPHCLTLRGORQSI
 IVAASRSEKEMWEDIQMTDLAEKSNPTPELLASSPDNKSPDEATADOSEDD
 LSASRTSLERQAPHRGNTWVHVCHRTSVSWDFSTAVENQLSGNLLRKFKNNGQ
 KLVVFTYFETKSHQDSPLASLLPLGSLTIPSENIHKDHYVFKLHFKRSHVY
 FRASEYTFEKKWMEVIRSATSSRAHLISLKESHLY"
 BASE COUNT 982 a 1083 c 997 g 840 t
 ORIGIN
 Query Match 82.1%; Score 631.4; DB 10; Length 3902;
 Best Local Similarity 89.7%; Pred. No. 2.2e-142;
 Matches 690; Conservative 0; Mismatches 76; Indels 3; Gaps 1;
 QY 4 CTTCTGTTTTCAGAGCACAGTGGACAAAGAGGAGCGCATGCGGAGCACTGAAAAGTC 63
 Db 1057 CTTCTGTTTTCAGAGCACAGTGGACAAAGAGGATCCATGCCCCAAGCCTTGAAGATC 1116
 QY 64 TCATATTCGGGAATTTTGAACCTTTTGCAAAATTTTCATACTAATTTTCTCAAGGAAATG 123
 Db 1117 TCATATTCGGGAATTTTGAACCTTTTGCAAAATTTTCTCAAGGAAATG 1176
 QY 124 AGCAACGACTTGCCTGTGGGAGGCGCTCAATGCCCAATCA--GAGATTACCAA 180
 Db 1177 AGCAACGACTTGCCTGTGGGAGGCGCTCAATGCCCAATCAAGGATGAGGATACCAA 1236
 QY 181 GAATCGGGGATCTGCTGAAGAACAATTCAGGCGCATGAAGCACTGCGGCTCACCTGT 240
 Db 1237 GAATCGGGGATCTGCTGAAGAACAATTCAGGCGCATGAAGCACTGCGGCGGCTCACCTGT 1296
 QY 241 GGAACACAGCAGCGCCTTGGAGGCCCTTGGAGATGGAATCAAGAGCTCCCGGCGGCTGG 300
 Db 1297 GGAACACAGTGGAGGCCCTTGGAGGCCCTTGGAGATGGAATCAAGAGCTCCCGGCGGCTGG 1356
 QY 301 AGAATCTTCGAGACATTTGAGCTGCAGAGGTTGTGTACCTACCGCTCAACACTTCC 360
 Db 1357 AATCTTCTGCGAGACTTCGAGCTGCAGAGGTTGTGTACCTACCGCTCAACACTTCC 1416
 QY 361 TCCTCGGCGCACTGCACCGGCTCATGCACACTACAGAGGTTCTCTGGAGCGGCTGTGCAAC 420
 Db 1417 TCCTCGGCGGCTGCACCGGCTCATGCACACTATAGCATGTCTCTGGAGAGGCTGTGCAAGC 1476
 QY 421 ACCACCCCGGAGCCACCGCACTTTCAGGAGCTGCCGAGCGCTTTTGGCAGAGATCAAGG 480

Db	Accession	Definition	Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Remark	Comment
Db 1477	ACCACCCACCAAAACACCCGACTTCAGGAGCTGCAGAGCTGCCTGGCGGAGATCACAG	1536											
QY 481	AGATGGTGGCAGCTCCACCGTACGATGATCAAGATGAGATTTCCAGAGCTTCACAG	540											
Db 1537	AGATGGTGGCAGCTCCACCGTACGATGATCAAGATGAGATTTCCAGAGCTTCACAG	1596											
QY 541	AACTCAAGAAAGATTGATTGGCATTGACAATCTTGTGTTCGCGGAAAGGAGTTTCATCC	600											
Db 1597	AGCTCAAGAAAGATTGATTGGCATTGACAATCTTGTGACCCCAAGGAGTTTCATCC	1656											
QY 601	GTCTGGGAGCTTCAGCAAGCTCTCGGGAAAGGGGCTCCAGCAGCGCATGTTTCTCTGT	660											
Db 1657	GCCTGGGAGCTTCAGCAAGCTCTCGGGAAAGGGGCTCCAGCAGCGCATGTTTCTCTGT	1716											
QY 661	TCAACGAGCTCTGCTATACAGCGCGGGGCTCACGGCTCCCAATCAGTTTAAAGTCC	720											
Db 1717	TCACGATGTTCTGCTGTATACCGCGGGGCTCACGGCTCCCAATCAGTTTAAAGTCC	1776											
QY 721	ACGGCAGCTCCCGCTCTATGGCATGACGATTGAGGAGGAGCGAAGACGA 769												
Db 1777	AGGACAGCTCCCACTCTATGGCATGACGATTGAGGAGGAGCGAAGACGA 1825												
RESULT 3													
LOCUS	BC004009	Mus musculus, clone IMAGE:3493093, mRNA, partial cds.											
DEFINITION	BC004009												
ACCESSION	BC004009.1	GI:13278387											
VERSION													
KEYWORDS		house mouse.											
SOURCE		house mouse.											
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.											
REFERENCE	1	(Bases 1 to 2632)											
AUTHORS		Strausberg, R.											
TITLE		Direct Submission											
JOURNAL		Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA											
REMARK		NIH-MGC Project URL: http://mgc.nci.nih.gov											
COMMENT		Contact: MGC help desk Email: cgapbs-re@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life technologies, Inc. DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcldpaxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.											

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 8 Row: p Column: 11
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

Location/Qualifiers
 1. .2632
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="CZECH II"
 /clone="IMAGE:3493093"
 /tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
 /clone_lib="NCI_CGAP_Lu29"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

QY 630 AAGGGCTCCAGCAGCGCATGTTCTTCCTGTTCAAGCAGCTCCTGCTATACAGCAGCCGG 689
 Db 241 AAGGGCTCCAGCAGCGCATGTTCTTCCTGTTCAAGCAGCTCCTGCTATACAGCAGCCGG 300
 QY 690 GGGCTGACGGCTCCCAATCAGTTTAAAGTCCAGGGCAGCTCCGCTCATGGCATGAGC 749
 Db 301 GGAAGTACGGCATCTAATCAGTTTAAAGTCCAGGGCAGCTCCCACTCATGGCATGAGC 360
 QY 750 ATTGAGGAGGAGGAGGAGCA 769
 Db 361 ATCGAGGAGGAGGAGGAGCA 380
 RESULT 5
 LOCUS AB018336 3997 bp mRNA linear PRI 16-JUN-1999
 DEFINITION Homo sapiens mRNA for KIAA0793 protein, complete cds.
 ACCESSION AB018336
 VERSION AB018336.1 GI:3882306
 KEYWORDS
 SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)
 Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,
 Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
 Prediction of the coding sequences of unidentified human genes. XI.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro
 DNA Res. 5 (5), 277-286 (1998)
 99087487
 2 (bases 1 to 3997)
 Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
 Direct Submission
 Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute,
 Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
 292-0812, Japan (E-mail:cdnaifn@kazusa.or.jp, Tel:+81-438-52-3913,
 Fax:+81-438-52-3914)
 FEATURES
 Location/Qualifiers
 1..3997
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hk05692"
 /sex="male"
 /tissue_type="brain"
 /clone_lib="pBluescriptII SK plus"
 /dev_stage="adult"
 1..3997
 /gene="KIAA0793"
 118..3282
 /gene="KIAA0793"
 /codon_start=1
 /product="KIAA0793 protein"
 /protein_id="BAA34513.1"
 /db_xref="GI:3882307"
 /translation="MGEIEGVRLQAGMLGAGTGVSTLEPGQTLPRMQEHL
 HLKVLNTWEIFDIEKQDQGVLLTVWRKLNLEVDYFGMEFQNTQYTWLEPM
 RPIHQIRRNKNVLRVLAFFPPDPQGLQEEYTRFLALQKRLDLLEERTCADTTA
 ALLSHLSQDEIGDYDETLDREHLKNVEYLPQOQHLKLEKILEFHOKHVGQTPAESDFQ
 VLETKARKLEMGIRFHMASDREGTKIQLAVSHMGLVFQGTTKINTFNWSVKRLSF
 RKRLIKLHPVHGYQDTLEFLGSRDECKNFWKICVEYHTFFLLDQPKPKAKAVF
 FSRGSSFRYSGRTOKQVDYFKDSGMKRIPIYERHSKTHTSVRALTADLPKQISFPE
 GLRTPASSSANAFYLSPSLTPSGELPEFDDSSSLSDTPQVSYVKSPPAAERSGAVA
 GQDTPPAQGLPPALQPGSLSPSSPSLSPSLSPAFQVLPAPAEQSSPL
 LSPVSDAGGAGMCEEPRHRRVPADEAYFIVKEILLATERTYLDKLEYIVWFRSAVF
 KEDAMPATLTLFSNIDPIYEFHRRGLREVEQRLALWEGPSKAHTKSHQRIIGILL
 RNMRLKEFTSYFQHDVLTLEKATKCKLEAVYKEFELOKVCYLPPLATFLKPI
 ORLLHRYLLRLGCHSPGHHDYADCHDALKAITEVTTTLQHLIRLENLQKLELQ
 RDLVGIENLAPGRFIREGCKLTKGLQRMFFLFSMDLLYTSKGVAGTSFRRIR
 GLPLQGLMVSQENSWPHCFITYAAQKTVVAASRTLEKRWMLDLNSAIQAQKS
 GGDTPALPGRVTCTPRSPNEVSLQESDADARGVRSLSLEGHQHQRANTTMHVCVY

RNTSVSRADHSAAVENQLSGYLLRRKFNHGWQKLVWVTFNFCFLFFYKTHDDYPLAS
 LPLGYSVSIPIREADGIHKDYVKLOKSHVYFFRAESKYITPERMVEVIOGASSAGR
 APSIVQDGPQPPSSGLEGMVRGKEE"
 BASE COUNT 998 a 1132 c 1068 g 799 t
 ORIGIN
 Query Match 36.5%; Score 280.8; DB 9; Length 3997;
 Best Local Similarity 61.1%; Pred. No. 2.1e-57;
 Matches 472; Conservative 0; Mismatches 297; Indels 3; Gaps 1;
 QY 1 TCACCTCGTGGTTTCAGAGCAGTGAAGAGAGGAGCGCATGCGGGAAGCACTGAAGA 60
 Db 1787 TTACCGTGTGGTTCCGACGCGCATGCTGAAGGAGGAGCGCATGCTGCGACTCATGA 1846
 QY 61 GTCCTCATATTCGGAATTTTGACCTTTGCACAAATTTTCATACTAATTTTCTCAAGAA 120
 Db 1847 CGTGTCTTCTTCCAACATCGATCCCATCTATAGATTTCCACAGAGGCTTCTCTCGCGAGG 1906
 QY 121 TTGAGCAAGCAGCTTGCCCTGTGGAAAGGCGCTCAATGCCCAATCAGAG---ATTACC 177
 Db 1907 TGGAGCAGAGGCTGGCAGCTCTGGAAAGGCGCTCCAAAGCCACACAAAGGCGAGTCATC 1966
 QY 178 AAGAATCGCGGATGTCATGCTGAAGAACAATTCAGGGCATGAAGCACTGGCGGCTCACC 237
 Db 1967 AACGAATCGGGACATCTCTGCTCAGGAACATGCGCCAGTTAAAGGAGTTTACCAGCTACT 2026
 QY 238 TGTGGAAGCACAGCAGGAGGCTTGGAGGCGCTTGAGAGTGAATCAAGAGCTCCCGCGGC 297
 Db 2027 TCAAGACATGACGAGGCTTGAAGAACTTGGAAGAGGCTACCAACCGCTGTGAAGAAGT 2086
 QY 298 TGGAGAACTTCTGACAGAGACTTTGAGCTGCAGAGGTGTGTTACCTACCGCTCAACACCT 357
 Db 2087 TGGAGCAGGTACAAAGGAGTTTGAGCTGCAGAAAGGTCTGCTACTTGCCTCTCAACACGT 2146
 QY 358 TCTCTCTGGGCACTGCACCGGCTCATGCACTACAGAGGTCTCTGGAGCGGCTGTGCA 417
 Db 2147 TCTGTCTGAAGCCCATCCAGCGGCTGCTGCACTACCGGCTGTGCTGCGCGGCTATGG 2206
 QY 418 AACACCAACCGCGGAGCGACGCGGCTTCAGGAGCTGCGGAGCGCTTTGGCAGAGATCA 477
 Db 2207 GACATTACAGCCCGGCGGACCATGACTACGCTGACTGCCATGACGCCCTGAAAGCCATCA 2266
 QY 478 CGGAGATGGTGACACACTCCACGGTACGATGATGAAGTGGAGAAATTTCCAGAAGCTGC 537
 Db 2267 CAGAGGTGACCAACACACTACAGCAGATCTCATCCGGCTGGAGAACTTCACAGAAGCTAA 2326
 QY 538 AGAACTCAAGAAAGATTTGATTGGCATTGACAATCTGTGGTTCCGGGAGGAGGTTC 597
 Db 2327 CGGAGCTGACGGGACCTGGTGGGCATAGAGAACTCATTTGCTCTGCGCAGGAGTTCA 2386
 QY 598 TCCGCTGGGAGCGCTCAGCAAGCTCTCGGGGAAGGCGTCCAGCAGCGCATCTCTTCC 657
 Db 2387 TCGTGAGGGCTGCTTCAAGCTCACCAGAAAGGCGCTGCGAGCAGAGATGTTTTTTC 2446
 QY 658 TGTTCACGACGCTCTCTGCTATACACAGCGGGGCGGTGACGGGCTCCAACTCAGTTTAAAG 717
 Db 2447 TGTTCACATATGTTGCTGTACAAAGCAAGGAGTTTCAGGAGCAGCCACTTCCGGA 2506
 QY 718 TCCACGGGAGCTCCCGCTCTATGCGATGACGATTGAGGAGCGGAGAGCA 769
 Db 2507 TCCGGGCGCTCTCTTCCCTCCCAAGGCGATGCTGTGTGAAGAAAGTATAACGA 2558
 RESULT 6
 LOCUS BC009153 3719 bp mRNA linear ROD 07-AUG-2002
 DEFINITION Mus musculus, similar to KIAA0793 gene product, clone MGC:6304
 IMAGE:2655209, mRNA, complete cds.
 ACCESSION BC009153
 VERSION BC009153.1 GI:14318718
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus

only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sv: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep>
 This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
 RP11-11L124 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-11L124. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
 The true left end of clone RP11-295B17 is at 104579 in this sequence. The true right end of clone RP11-261P24 is at 2000 in this sequence.

FEATURES

Source	Location/Qualifiers
	1..106578
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="13"
	/map="q31.3-32.3"
	/clone="RP11-11L124"
	/clone_lib="RPI1-11.1"
	79533
misc_feature	/note="Tandem repeat. Forced join. Gap size estimated to be approximately 800bp by restriction digest data"
BASE COUNT	28669 a 25979 c 25140 g 26790 t
ORIGIN	
Query Match	31.3%; Score 241; DB 9; Length 106578;
Best Local Similarity	100.0%; Pred. NO. 1.le-47;
Matches 241;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 219	AAGCACTGGCGCTCACCTGTGGAGACAGCGAGCGCTTGGAGCCCTGGAGAAATGGA 278
Db 37298	AAGCACTGGCGCTCACCTGTGGAGACAGCGAGCGCTTGGAGCCCTGGAGAAATGGA 37357
QY 279	ATCAAGAGCTCCGCGCGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGAGGTGTGT 338
Db 37358	ATCAAGAGCTCCGCGCGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGAGGTGTGT 37417
QY 339	TACCTACCGCTCAACACCTTCTCTGCGGCACTGCACGGCTCATGCACTACAGCAG 398
Db 37418	TACCTACCGCTCAACACCTTCTCTGCGGCACTGCACGGCTCATGCACTACAGCAG 37477
QY 399	GTCTCGAGCGCTGTGCAAAACACACCGCGAGCGCGACCTTCAGGAGCTCCCGA 458
Db 37478	GTCTCGAGCGCTGTGCAAAACACACCGCGAGCGCGACCTTCAGGAGCTCCCGA 37537
QY 459	G 459
Db 37538	G 37538
RESULT 8	
AC094777	
LOCUS	AC094777 224310 bp DNA linear HTG 24-AUG-2002
DEFINITION	Rattus norvegicus clone CH230-4N8, *** SEQUENCING IN PROGRESS ***
ACCESSION	AC094777
VERSION	AC094777.6 GI:22164997

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 224310)

Muzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Delvalle, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Din, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flag, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Gant, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, H., Loulseg, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., Mcleod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwako, O., Okwuonu, G., Otlorunpogon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaj, D., Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

2 (bases 1 to 224310)

Worley, K.C.

REFERENCE

AUTHORS

TITLE

JOURNAL

3 (bases 1 to 224310)

Rat Genome Sequencing Consortium.

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Aug 9, 2002 this sequence version replaced gi:22094284.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBIZ

Center clone name: CH230-4N8

----- Summary Statistics

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 172428 bases at least Q40

Consensus quality: 179331 bases at least Q30

Consensus quality: 184096 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 65 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1
1011: contig of 1011 bp in length
1012 1111: gap of unknown length
1112 2458: contig of 1347 bp in length
2459 2558: gap of unknown length
2559 3837: contig of 1278 bp in length
3837 3937: gap of unknown length
3937 5467: contig of 1531 bp in length
5468 5567: gap of unknown length
5568 6863: contig of 1296 bp in length
6864 6963: gap of unknown length
6964 8013: contig of 1050 bp in length
8014 8113: gap of unknown length
8114 9431: contig of 1318 bp in length
9432 9531: gap of unknown length
9532 10970: contig of 1439 bp in length
10971 11070: gap of unknown length
11071 12178: contig of 1108 bp in length
12179 12278: gap of unknown length
12279 13604: contig of 1326 bp in length
13605 13704: gap of unknown length
13705 15160: contig of 1456 bp in length
15161 15260: gap of unknown length
15261 16520: contig of 1260 bp in length
16521 16620: gap of unknown length
16621 18279: contig of 1659 bp in length
18280 18379: gap of unknown length
18380 19562: contig of 1183 bp in length
19563 19662: gap of unknown length
19663 20728: contig of 1066 bp in length
20729 20828: gap of unknown length
20829 22059: contig of 1231 bp in length
22060 22159: gap of unknown length
22160 24977: contig of 2818 bp in length
24978 25077: gap of unknown length
25078 26432: contig of 1355 bp in length
26433 26532: gap of unknown length
26533 27664: contig of 1132 bp in length
27665 27764: gap of unknown length
27765 29834: contig of 2070 bp in length
29835 29934: gap of unknown length
29935 32076: contig of 2142 bp in length
32077 32176: gap of unknown length
32177 33953: contig of 1777 bp in length
33954 34053: gap of unknown length
34054 36346: contig of 2293 bp in length
36347 36446: gap of unknown length
36447 38429: contig of 1983 bp in length
38430 38529: gap of unknown length
38530 41113: contig of 2384 bp in length
41114 41213: gap of unknown length
41214 43096: contig of 1883 bp in length

* 43097 43196: gap of unknown length
43197 44735: contig of 1539 bp in length
44736 44835: gap of unknown length
44836 46649: contig of 1814 bp in length
46650 46749: gap of unknown length
46750 48470: contig of 1721 bp in length
48471 48570: gap of unknown length
48571 50495: contig of 1925 bp in length
50496 50595: gap of unknown length
50596 52568: contig of 1973 bp in length
52569 52668: gap of unknown length
52669 54428: contig of 1760 bp in length
54429 54528: gap of unknown length
54530 58301: contig of 3773 bp in length
58302 58401: gap of unknown length
58402 62471: contig of 4070 bp in length
62472 62571: gap of unknown length
62572 64513: contig of 1942 bp in length
64514 64614: gap of unknown length
64615 66942: contig of 2329 bp in length
66943 67042: gap of unknown length
67043 69741: contig of 2699 bp in length
69742 69841: gap of unknown length
69842 72553: contig of 2712 bp in length
72554 72653: gap of unknown length
72654 74402: contig of 1749 bp in length
74403 74502: gap of unknown length
74503 77297: contig of 2795 bp in length
77298 77397: gap of unknown length
77398 81383: contig of 3986 bp in length
81384 81484: gap of unknown length
81485 83907: contig of 2424 bp in length
83908 84007: gap of unknown length
84008 86993: contig of 2986 bp in length
86994 87093: gap of unknown length
87094 89399: contig of 2306 bp in length
89400 89499: gap of unknown length
89500 93060: contig of 3561 bp in length
93061 93160: gap of unknown length
93161 97688: contig of 4528 bp in length
97689 97788: gap of unknown length
97789 102376: contig of 4588 bp in length
102377 102476: gap of unknown length
102477 106253: contig of 3777 bp in length
106254 106353: gap of unknown length
106354 109283: contig of 2930 bp in length
109284 109383: gap of unknown length
109384 112196: contig of 2813 bp in length
112197 112296: gap of unknown length

Query Match 26.6%; Score 204.2; DB 2; Length 224310;
Best Local Similarity 90.5%; Pred. No. 9.8e-39;

Matches 218; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 219 AAGCACCTGGCGGCTCACCTGTGGAAGCACACGAGGCGCTTGAGGCCCTGGAGATGGA 278
|||||

Db 79335 AAGCACCTGGCGGCGGCTCACCTGTGGAAGCACACGAGGCGCTTGAGGCCCTGGAGACTCC 79394
|||||

QY 279 ATCAAGAGCTCCGGCGGCTGGAGAACTTTCGAGAGACTTTGAGCTGCAGAGGTGTGT 338
|||||

Db 79395 ATCAAGAGGCTCCGGCGGCTGGAGAACTTTCGAGAGACTTTGAGCTGCAGAGGTGTGT 79454
|||||

QY 339 TACCTACCGCTCAACACCTTCTCTGCGGCGCTGACCGGCTCATGCTACAGAGCAG 398
|||||

Db 79455 TACCTACCGCTCAACACCTTCTCTGCGGCGCTGACCGGCTCATGCTACAGAGCAG 79514
|||||

QY 399 GTCTGGAGCGCTGTGCAACACACCGCGGCGGAGCGGCGGCTTTCAGGACTGCCGA 458
|||||

Db 79515 GTCTGGAGCGCTGTGCAACACACCGCGGCGGAGCGGCGGCTTTCAGGACTGCCGA 79574
|||||

QY 459 G 459
|

Db 79575 G 79575

RESULT 9	AC122885	221502 bp	DNA	linear	HTG 23-JUN-2002
LOCUS	Mus musculus chromosome UNK clone RP23-90H14, WORKING DRAFT				
DEFINITION	SEQUENCE, 8 unordered pieces.				
ACCESSION	AC122885				
VERSION	AC122885.2	GI:21539806			
KEYWORDS	HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	McPherson,J.D. and Waterston,R.H.				
JOURNAL	The sequence of Mus musculus clone				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 221502)				
TITLE	McPherson,J.D. and Waterston,R.H.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park				
AUTHORS	Parkway, St. Louis, MO 63108, USA				
TITLE	3 (bases 1 to 221502)				
JOURNAL	McPherson,J.D. and Waterston,R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park				
TITLE	Parkway, St. Louis, MO 63108, USA				
JOURNAL	On Jun 23, 2002 this sequence version replaced gi:21218511.				
COMMENT	----- Genome Center ----- Center: Washington University Genome Sequencing Center Center code: WUGSC Web site: http://genome.wustl.edu/gsc/index.shtml Contact: submissions@watson.wustl.edu ----- Project Information ----- Center project name: M_BA0090H14 ----- Summary Statistics ----- Sequencing vector: M13; 0% Sequencing method: plasmid; 100% Chemistry: Dye-terminator Big Dye; 100% of reads Assembly: Phrap; version 0.990319 Consensus quality: 216976 bases at least Q40 Consensus quality: 217754 bases at least Q30 Consensus quality: 218341 bases at least Q20 Insert size: 190000; agarose-fp Insert size: 221868; sum-of-contigs Quality coverage: 15.60 in Q20 bases; agarose-fp Quality coverage: 12.41 in Q20 bases; sum-of-contigs ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 8 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 8441: contig of 8441 bp in length * 8442 8541: gap of unknown length * 8542 27277: contig of 18736 bp in length * 27278 27377: gap of unknown length * 27378 42339: contig of 14962 bp in length * 42340 42439: gap of unknown length * 42440 71538: contig of 29099 bp in length * 71539 71638: gap of unknown length * 71639 99587: contig of 27949 bp in length * 99588 99687: gap of unknown length * 99688 134048: contig of 34361 bp in length * 134049 134148: gap of unknown length * 134149 221349: contig of 87201 bp in length				

Fax: (650) 320-5801
Email: olivier@hgc.stanford.edu
Primer A: AGAGATGCTTTATCTCCNCCG
Primer B: ACGTAGGATGATCAAGATGGAG
STS size: 290
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifrag Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul
Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3
Finished human sequence in NCBI. STSs designed and developed at the Stanford Human Genome Center.
Location/Qualifiers
1. .393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="13"
/clone_lib="Human"
12. .301
12. .34
primer_bind complement(279..301)
BASE COUNT 80 a 105 c 93 g 115 t
ORIGIN
Query Match 21.1%; Score 162.4; DB 11; Length 393;
Best Local Similarity 95.9%; Pred. No. 8.7e-29;
Matches 188; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY 409 GGCTGTGCAAAACA-CCACCCGCGAGCCGACCTTCAGGACTGCGGACCGCTTG 467
|||||
Db 393 GGCTGTGCAAAACACCACCCGCGAGCCGACCTTCAGGACTGCGGACCGCTTG 334
|||||
QY 468 GCAGATGATCAGGAGATGTTGGACAGC-TCCACGGTACGATGATCAAGATGGAGATTT 526
|||||
Db 333 GCAGATGATCAGGAGATGTTGGACAGCTTCCACGGTACGATGATCAAGATGGAGATTT 274
|||||
QY 527 CCAGAGCTGCAGCACTCAAGAAAGATTTGATGGCATTGACATCTTGTGTTCCGGG 586
|||||
Db 273 CCAGAGCTGCAGCACTCAAGAAAGATTTGATGGCATTGACATCTTGTGTTCCGGG 214
|||||
QY 587 AAGGGAGTTTCATCCG 602
|||||
Db 213 AAGGGTAAGCAGCAGT 198
|||||
RESULT 11
AC022669
LOCUS AC022669 Homo sapiens chromosome 13 clone RP11-96B23 map 13, WORKING DRAFT
DEFINITION AC022669
ACCESSION AC022669
VERSION AC022669.4 GI:9966291
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152053)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Homo sapiens chromosome 13, clone RP11-96B23
Unpublished
2 (bases 1 to 152053)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A., and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 152053)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
Collins,S., Collamore,A., Cooke,P., Dearellano,K., Dewar,K.,
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
Iliev,I., Johnson,R., Jones,C., Karatas,A., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Sougniez,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,W.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:7229803.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I4938
Center clone name: 96_B_23
----- Summary Statistics
Sequencing vector: M13; M7815; 93% of reads
Sequencing vector: Plasmid; n/a; %0.f% of reads
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of
reads
Assembly program: Phrap; version 0.960731
Consensus quality: 136451 bases at least Q40
Consensus quality: 143612 bases at least Q30
Consensus quality: 146365 bases at least Q20
Insert size: 158000; agarose-fp
Insert size: 148853; sum-of-contents
Quality coverage: 3.4 in Q20 bases; agarose-fp

Quality cov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 784: contig of 784 bp in length
* 785 884: gap of 100 bp
* 885 1956: contig of 1072 bp in length
* 1957 2056: gap of 100 bp
* 2057 3113: contig of 1057 bp in length
* 3114 3213: gap of 100 bp
* 3214 4817: contig of 1604 bp in length
* 4818 4917: gap of 100 bp
* 4918 6582: contig of 1665 bp in length
* 6583 6682: gap of 100 bp
* 6683 8059: contig of 1377 bp in length
* 8060 8159: gap of 100 bp
* 8160 22369: contig of 14210 bp in length
* 22370 22469: gap of 100 bp
* 22470 24974: contig of 2505 bp in length
* 24975 25074: gap of 100 bp
* 25075 27206: contig of 2132 bp in length
* 27207 27306: gap of 100 bp
* 27307 29420: contig of 2114 bp in length
* 29421 29520: gap of 100 bp
* 29521 32000: contig of 2480 bp in length
* 32001 32100: gap of 100 bp
* 32101 36330: contig of 4230 bp in length
* 36331 36430: gap of 100 bp
* 36431 39067: contig of 2637 bp in length
* 39068 39167: gap of 100 bp
* 39168 42564: contig of 3397 bp in length
* 42565 42664: gap of 100 bp
* 42665 46664: contig of 4000 bp in length
* 46665 46764: gap of 100 bp
* 46765 51047: contig of 4283 bp in length
* 51048 51147: gap of 100 bp
* 51148 55269: contig of 4122 bp in length
* 55270 55369: gap of 100 bp
* 55370 58831: contig of 3462 bp in length
* 58832 58931: gap of 100 bp
* 58932 64077: contig of 5146 bp in length
* 64078 64177: gap of 100 bp
* 64178 68808: contig of 4631 bp in length
* 68809 68908: gap of 100 bp
* 68909 74170: contig of 5262 bp in length
* 74171 74270: gap of 100 bp
* 74271 79296: contig of 5026 bp in length
* 79297 79396: gap of 100 bp
* 79397 84724: contig of 5328 bp in length
* 84725 84824: gap of 100 bp
* 84825 89713: contig of 4889 bp in length
* 89714 89813: gap of 100 bp
* 89814 94721: contig of 4908 bp in length
* 94722 94821: gap of 100 bp
* 94822 100196: contig of 5375 bp in length
* 100197 100296: gap of 100 bp
* 100297 107015: contig of 6719 bp in length
* 107016 107115: gap of 100 bp
* 107116 114507: contig of 7392 bp in length
* 114508 114607: gap of 100 bp
* 114608 122671: contig of 8064 bp in length
* 122672 122771: gap of 100 bp
* 122772 128052: contig of 5281 bp in length
* 128053 128152: gap of 100 bp
* 128153 138583: contig of 10431 bp in length
* 138584 138683: gap of 100 bp
* 138684 150248: contig of 11565 bp in length
* 150249 150348: gap of 100 bp

* 150349 152053: contig of 1705 bp in length.
FEATURES
Source
Location/Qualifiers
1. .152053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13"
/clone="RP11-96B23"
/clone_lib="RPC1-11 Human Male BAC"
1. .784
/note="assembly_fragment"
clone_end:Sp6
vector_side:left
885. .1956
/note="assembly_fragment"
2057. .3113
/note="assembly_fragment"
3214. .4817
/note="assembly_fragment"
4918. .6582
/note="assembly_fragment"
6683. .8059
/note="assembly_fragment"
8160. .22369
/note="assembly_fragment"
22470. .24974
/note="assembly_fragment"
25075. .27206
/note="assembly_fragment"
27307. .29420
/note="assembly_fragment"
29521. .32000
/note="assembly_fragment"
32101. .36330
/note="assembly_fragment"
36431. .39067
/note="assembly_fragment"
39168. .42564
/note="assembly_fragment"
42665. .46664
/note="assembly_fragment"
46765. .51047
/note="assembly_fragment"
Query Match 17.8%; Score 136.8; DB 2; Length 152053;
Best Local Similarity 98.6%; Pred. No. 2.1e-22;
Matches 138; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 8 GTGGTTTCAGAGCACAGTCAGCAAGAGAGGACGCCATGCCGGAAGTCTCAT 67
|||||
Db 103907 GTGGTTTCAGAGCACAGTCAGCAAGAGAGGACGCCATGCCGGAAGTCTCAT 103966
QY 68 ATTCCCGAATTTTGACCTTTGCACAAATTTTCATCTAATTTTCTCAAGAAATTCAGCA 127
|||||
Db 103967 ATTCCCGAATTTTGACCTTTGCACAAATTTTCATCTAATTTTCTCAAGAAATTCAGCA 104026
QY 128 ACGACTTGCCCTGTGGGAAG 147
|||||
Db 104027 ACGACTTGCCCTGTGGTGAG 104046
RESULT 12
AX261839/c
LOCUS AX261839 426 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 1490 from Patent WO0173027.
ACCESSION AX261839
VERSION AX261839.1 GI:16510806
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1

AUTHORS Meagher, M.J., Xu, J. and King, G.E.
TITLE Compositions and methods for therapy and diagnosis of colon cancer
PATENT WO 0173027-A 1490 04-OCT-2001;
CORIXA CORPORATION (US)

FEATURES
source
1. 426

BASE COUNT 89 a 109 c 109 g 119 t
ORIGIN /organism="Homo sapiens"
/db_xref="taxon:9606"

Query Match 15.7%; Score 120.8; DB 6; Length 426;
Best Local Similarity 60.2%; Pred. No. 1.1e-18;
Matches 219; Conservative 0; Mismatches 142; Indels 3; Gaps 1;

QY 1 TCACCTCGTGGTTTCAGGACACAGTCAGCAAGAGAGAGCCATGCGCGGAGACACTGCAAAA 60
DB 365 TTACCGTGTGGTTCGCCAGCGAGTGGTGAAGGAGGAGCCATGCGCTGCGACTCTGATGA 306
QY 61 GTCTCATATTCGCCGAATTTTGAACCTTTCACAAAATTTCTATACTAATTTTCACAGGAAA 120
DB 305 CGTCTCTCTTCCACATCGATCCATCTATGAGTTCCACAGAGGCTTCTTCCGCGGAGG 246
QY 121 TTGAGCAGCACTTCCCTGTGGGAGCGCGCTCAAAATGCCAAATCAGAG---ATTACC 177
DB 245 TGGAGCAGAGGCTGGCACTCTGGGAAGGCGCCTCCAAAGCCACACAAAAGCGAGTCATC 186
QY 178 AAGAAATCGGCATGTCATGCTGAAGACATTCAGGCGCATGAAGCACTGGCGGGTCACC 237
DB 185 ACGAAATCGGAGACATCTCGCTCAGAGATCGCGCAGTTAAAGGAGTTTACCACTACT 126
QY 238 TGTGAAGCAGCAGCGAGGCGCTTGGAGCGCCTTGGGAATGGAATCAAGAGCTCCCGCGGC 297
DB 125 TCCAAAGACATGACGAGATCCTTAACAGACTGGAAGAGCTACCAACGCTGTAGAAGT 56
QY 298 TGGAGAACTTTCGACAGACTTTGAGCTGCAGAGAGTGTTTACCTACCGCTCAACACT 357
DB 65 TGGAGCAGCTGTACAGAGAGTTTGAAGTGCAGAGAGTCTGCTACTTGCCTCTCAACACT 6
QY 358 TCCT 361
DB 5 TCCT 2

RESULT 13
AC101725/c
LOCUS AC101725 190588 bp DNA linear HTG 21-AUG-2002
DEFINITION Mus musculus clone RP23-321M11, WORKING DRAFT SEQUENCE, 28
unordered pieces.

ACCESSION AC101725
VERSION AC101725.2 GI:22381046
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 190588)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus, clone RP23-321M11

Unpublished

2 (bases 1 to 190588)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 190588)

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats 21, 2002 this sequence version replaced gi:17060500.

On all repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L17278

Center clone name: 321_M11

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 182743 bases at least Q40

Consensus quality: 185393 bases at least Q30

Consensus quality: 186552 bases at least Q20

Insert size: 206000; agarose-fp

Insert size: 187888; sum-of-ctnigs

Quality coverage: 5.7 in Q20 bases; agarose-fp

Quality coverage: 6.3 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 698: contig of 698 bp in length

* 699 798: gap of 100 bp

* 799 1437: contig of 639 bp in length

* 1438 1537: gap of 100 bp

* 1538 2255: contig of 718 bp in length

* 2256 2355: gap of 100 bp

```
* 2356 3033: contig of 678 bp in length
* 3034 3133: gap of 100 bp
* 3134 3783: contig of 650 bp in length
* 3784 3883: gap of 100 bp
* 3884 4559: contig of 676 bp in length
* 4560 4659: gap of 100 bp
* 4660 5309: contig of 650 bp in length
* 5310 5409: gap of 100 bp
* 5410 6053: contig of 644 bp in length
* 6054 6153: gap of 100 bp
* 6154 7009: contig of 856 bp in length
* 7010 7109: gap of 100 bp
* 7110 7747: contig of 638 bp in length
* 7748 7847: gap of 100 bp
* 7848 8125: contig of 278 bp in length
* 8126 8225: gap of 100 bp
* 8226 8956: contig of 731 bp in length
* 8957 9056: gap of 100 bp
* 9057 10142: contig of 1086 bp in length
* 10143 10242: gap of 100 bp
* 10243 11404: contig of 1162 bp in length
* 11405 11504: gap of 100 bp
* 11505 12593: contig of 1089 bp in length
* 12594 12693: gap of 100 bp
* 12694 14681: contig of 1988 bp in length
* 14682 14781: gap of 100 bp
* 14782 17534: contig of 2753 bp in length
* 17535 17634: gap of 100 bp
* 17635 21928: contig of 4294 bp in length
* 21929 22028: gap of 100 bp
* 22029 29432: contig of 7404 bp in length
* 29433 29532: gap of 100 bp
* 29533 35637: contig of 6105 bp in length
* 35638 35737: gap of 100 bp
* 35738 84581: contig of 48844 bp in length
* 84582 84681: gap of 100 bp
* 84682 93581: contig of 8900 bp in length
* 93582 93681: gap of 100 bp
* 93682 101846: contig of 8165 bp in length
* 101847 101946: gap of 100 bp
* 101947 111400: contig of 9454 bp in length
* 111401 111500: gap of 100 bp
* 111501 131247: contig of 19747 bp in length
* 131248 131347: gap of 100 bp
* 131348 151032: contig of 19685 bp in length
* 151033 151132: gap of 100 bp
* 151133 184681: contig of 33549 bp in length
* 184682 184781: gap of 100 bp
* 184782 190588: contig of 5807 bp in length.
FEATURES
    source
        1..190588
            Location/Qualifiers
                organism="Mus musculus"
                db_xref="taxon:10090"
                clone="RP23-321M11"
                clone_lib="RPC1-23 Female Mouse BAC"
            1..698
                /note="assembly_fragment"
                799..1437
                /note="assembly_fragment"
                1538..2255
                /note="assembly_fragment"
                2356..3033
                /note="assembly_fragment"
                3134..3783
                /note="assembly_fragment"
                3884..4559
                /note="assembly_fragment"
                4660..5309
                /note="assembly_fragment"
                5410..6053
                /note="assembly_fragment"
                6154..7009
                /note="assembly_fragment"
```

```
misc_feature
    7110..7747
        /note="assembly_fragment"
misc_feature
    7848..8125
        /note="assembly_fragment"
misc_feature
    8226..8956
        /note="assembly_fragment"
misc_feature
    9057..10142
        /note="assembly_fragment"
misc_feature
    10243..11404
        /note="assembly_fragment"
misc_feature
    11505..12593
        /note="assembly_fragment"
misc_feature
    12694..14681
        /note="assembly_fragment"
misc_feature
    14782..17534
        /note="assembly_fragment"
misc_feature
    17635..21928
        /note="assembly_fragment"
misc_feature
    22029..29432
        /note="assembly_fragment"
misc_feature
    29533..35637
        /note="assembly_fragment"
misc_feature
    35738..84581
        /note="assembly_fragment"
misc_feature
    84682..93581
        /note="assembly_fragment"
misc_feature
    93682..101846
```

Query Match 15.5%; Score 119.2; DB 2; Length 190588;
Best Local Similarity 90.7%; Pred. No. 3.8e-18;
Matches 127; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
QY 8 GGGTTTCAGACGACGTCAGCAAGAGAGACCCATCGCGGAGCAGCTCAAGTCTCAT 67
    |||||
Db 185168 GTGGTTTCAGACGACGTCAGCAAGAGAGCTCCAGCCCGAGCTTGAAGAGTCTCAT 185109

QY 68 ATTCGCCGAATTTTGAACTTTGCACAAATTTCTACTATTTTCTCAAGAAATTCAGCA 127
    |||||
Db 185108 ATTCGCCAAATTTTGAACTTTGCACAAATTTCTACTATTTTCTCAAGAAATTCAGCA 185049

QY 128 ACGACTTGCCTGTGGGAG 147
    |||||
Db 185048 ACGACTTGCCTGTGGTGAG 185029
```

RESULT 14
AC122885/c

LOCUS AC122885 221502 bp DNA linear HTG 23-JUN-2002
DEFINITION Mus musculus chromosome UNK clone RP23-90H14, WORKING DRAFT
SEQUENCE, 8 unordered pieces.

ACCESSION AC122885
VERSION AC122885.2 GI:21539806
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 221502)
REFERENCE McPherson,J.D. and Waterston,R.H.

REFERENCE 2 (bases 1 to 221502)
JOURNAL The sequence of Mus musculus clone
AUTHORS Unpublished

REFERENCE 3 (bases 1 to 221502)
JOURNAL Direct Submission
AUTHORS McPherson,J.D. and Waterston,R.H.

REFERENCE 3 (bases 1 to 221502)
JOURNAL Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS Parkway, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 221502)
JOURNAL Direct Submission
AUTHORS McPherson,J.D. and Waterston,R.H.

REFERENCE 3 (bases 1 to 221502)
JOURNAL Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS Parkway, St. Louis, MO 63108, USA

COMMENT On Jun 23, 2002 this sequence version replaced gi:21218511.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BA090H14
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 216976 bases at least Q40
Consensus quality: 217754 bases at least Q30
Consensus quality: 218341 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 221868; sum-of-contigs
Quality coverage: 15.60 in Q20 bases; agarose-fp
Quality coverage: 12.41 in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 8441: contig of 8441 bp in length
* 8442 8541: gap of unknown length
* 8542 27277: contig of 18736 bp in length
* 27278 27377: gap of unknown length
* 27378 42339: contig of 14962 bp in length
* 42340 42439: gap of unknown length
* 42440 71538: contig of 29099 bp in length
* 71539 99587: gap of unknown length
* 99588 99687: contig of 27949 bp in length
* 99688 134048: gap of unknown length
* 134149 221349: contig of 34361 bp in length
* 221350 221449: gap of unknown length
* 221450 221502: contig of 87201 bp in length
* 221502 221502: contig of 53 bp in length.

```

FEATURES

```

Source
1..221502
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-90H14"
misc_feature
1..8441
/note="assembly_name:Contig59"
misc_feature
8542..27277
/note="assembly_name:Contig60"
misc_feature
27378..42339
/note="assembly_name:Contig61"
misc_feature
42440..71538
/note="assembly_name:Contig62"
misc_feature
71639..99587
/note="assembly_name:Contig63"
misc_feature
99688..134048
/note="assembly_name:Contig64"
misc_feature
134149..221349
/note="assembly_name:Contig65"
misc_feature
221450..221502
/note="assembly_name:Contig19"
BASE COUNT 61405 a 54522 c 50450 g 54373 t 752 others
ORIGIN

```

Query Match 15.5%; Score 119.2; DB 2; Length 221502;
 Best Local Similarity 90.7%; Pred. No. 3.9e-18;
 Matches 127; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

QY 8 GTGCTTTCAGAGCACAGTGTGAGCAAGAGGAGCGCCATCGCGAAGCACTGAAAGTCTCAT 67
|||||
Db 111066 GTGCTTTCAGAGCACAGTGTGAGCAAGAGGAGCTCCATCGCGAAGCCCTTGAGAGTCTCAT 111007
|||||

QY 68 ATTCGCCAAATTTTGAACCTTTTGACAAAATTCATACATAATTTTCTCAAGGAAATTTGAGCA 127
|||||
Db 111006 ATTCGCCAAATTTTGAACCTTTTGACAAAATTCATACATAATTTTCTCAAGGAAATTTGAGCA 110947
|||||

QY 128 ACGACTTGCCTGTGGGAAG 147
- |||||
Db 110946 ACGACTTGCCTGTGGTGAG 110927
|||||

```

RESULT 15

AC118779

LOCUS AC118779 149417 bp DNA linear HTG 18-JUL-2002
 DEFINITION Rattus norvegicus clone CH230-255J6, *** SEQUENCING IN PROGRESS
 ***, 61 unordered pieces.

AC118779 2 GI:21746687

VERSION HTG; HTGS-PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 149417)

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
 Homsli,F., Howard,S., Huber,J., Huiy,J., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Saverly,G.,
 Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
 Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

Direct Submission

JOURNAL

REFERENCE 2 (bases 1 to 149417)

AUTHORS

Worley,K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (21-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 149417)
Worley K.C.

Direct Submission

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20258202.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GUWA

Center Clone name: CH230-255J6

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 106401 bases at least Q40

Consensus quality: 113612 bases at least Q30

Consensus quality: 117819 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: this is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1127: contig of 1127 bp in length
1128 1227: gap of unknown length
1228 2235: contig of 1008 bp in length
2236 2335: gap of unknown length
2336 3792: contig of 1457 bp in length
3793 3892: gap of unknown length
3893 4977: contig of 1085 bp in length
4978 5077: gap of unknown length
5078 6106: contig of 1029 bp in length
6107 6206: gap of unknown length
6207 7617: contig of 1411 bp in length
7618 7717: gap of unknown length
7718 9305: contig of 1588 bp in length
9306 9405: gap of unknown length
9406 10632: contig of 1227 bp in length
10633 12158: contig of 1426 bp in length
12159 12258: gap of unknown length
12259 13281: contig of 1023 bp in length
13282 13381: gap of unknown length
13382 15064: contig of 1683 bp in length
15065 15164: gap of unknown length
15165 17718: contig of 2554 bp in length
17719 17818: gap of unknown length
17819 19363: contig of 1545 bp in length
19364 19463: gap of unknown length
19464 20526: contig of 1063 bp in length
20527 20626: gap of unknown length
20627 21864: contig of 1238 bp in length
21865 21964: gap of unknown length
21965 23838: contig of 1874 bp in length
23839 23938: gap of unknown length
23939 25034: contig of 1096 bp in length
25035 25134: gap of unknown length
25135 27370: contig of 2236 bp in length
27371 27470: gap of unknown length
28692: contig of 1222 bp in length

28792: gap of unknown length
30075: contig of 1283 bp in length
30175: gap of unknown length
31470: contig of 1295 bp in length
31570: gap of unknown length
32895: contig of 1325 bp in length
32995: gap of unknown length
34801: contig of 1806 bp in length
34901: gap of unknown length
36662: contig of 1761 bp in length
36762: gap of unknown length
37997: contig of 1235 bp in length
38097: gap of unknown length
40561: contig of 2464 bp in length
40661: gap of unknown length
42524: contig of 1863 bp in length
42624: gap of unknown length
44247: contig of 1623 bp in length
44347: gap of unknown length
46510: contig of 2163 bp in length
46610: gap of unknown length
48669: contig of 2059 bp in length
48769: gap of unknown length
50218: contig of 1449 bp in length
50318: gap of unknown length
52454: contig of 2136 bp in length
52554: gap of unknown length
53956: contig of 1402 bp in length
54056: gap of unknown length
56174: contig of 2118 bp in length
56274: gap of unknown length
58482: contig of 2208 bp in length
58582: gap of unknown length
60511: contig of 1929 bp in length
60611: gap of unknown length
62512: contig of 1901 bp in length
62612: gap of unknown length
65352: contig of 2740 bp in length
65452: gap of unknown length
67770: contig of 2318 bp in length
67870: gap of unknown length
70795: contig of 2925 bp in length
70895: gap of unknown length
73397: contig of 2502 bp in length
73497: gap of unknown length
75528: contig of 2031 bp in length
75628: gap of unknown length
78573: contig of 2945 bp in length
78673: gap of unknown length
80983: contig of 2210 bp in length
80984: gap of unknown length
83541: contig of 2558 bp in length
83641: gap of unknown length
86220: contig of 2579 bp in length
86320: gap of unknown length
89057: contig of 2737 bp in length
89157: gap of unknown length
92362: contig of 3205 bp in length
92462: gap of unknown length
95949: contig of 3487 bp in length
96049: gap of unknown length
100396: contig of 4347 bp in length
100496: gap of unknown length
104325: contig of 3829 bp in length
104425: gap of unknown length
107324: contig of 2899 bp in length
107424: gap of unknown length
110542: contig of 3118 bp in length
110642: gap of unknown length
113543: contig of 2901 bp in length

Query Match 14.5%; Score 111.8; DB 2; Length 149417;
Best Local Similarity 87.8%; Pred. No. 2.4e-16;

	Matches	122;	Conservative	0;	Mismatches	17;	Indels	0;	Gaps	0;
Qy	453	TGCGAGCGGCTTTGGCAGAGATCACGGAGATGGTGGCAGAGCTCCACGGTACGATGATC	512							
Db	28953	TCCTCAGCTGGGCTGGCAGAGATCACAGAGATGGTGGCAGAGCTCCACGGTACCATGATC	29012							
Qy	513	AAGATGGAGAAATTTCCAGAAGCTGCACGAACTCAAGAAAGATTTGATTGGCATTGACAAT	572							
Db	29013	AAGATGGAGAACTTCCAGAAGCTGCATGAGCTCAAGAAGGATCTGATTGGCATTGACAAT	29072							
Qy	573	CTTGTGGTTCGGGAAGGG	591							
Db	29073	CTTGTGATCCAGGAAGGG	29091							

Search completed: December 12, 2002, 06:06:55
Job time : 3622 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 02:37:49 ; Search time 306 seconds

(without alignments)
5659.434 Million cell updates/sec

Title: US-09-555-342b-1_COPY_1733_2501

Perfect score: 769

Sequence: 1 tcaactcgtggttcagacg.....attgaggagacgaagacga 769

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	769	100.0	3187	23	AAS64760 DNA encoding novel
2	769	100.0	3442	20	AAX79183 Human chondrocyte-
3	769	100.0	4687	21	AAA08582 Human cytoskeleton
4	762	99.1	3094	21	AAC98992 Human pancreatic c
5	120.8	15.7	426	23	AAS58814 cDNA #1490 encodin
6	106	13.8	1718	22	AAS27016 cDNA encoding nove
7	105.2	13.7	716	22	AAS27440 cDNA encoding nove
8	98	12.7	717	24	ABL90828 Human polynucleoti
9	97.2	12.6	1747	23	ABL03867 Drosophila melanog

C	10	97.2	12.6	4150	23	ABL03866	Drosophila melanog
	11	90.2	11.7	544	22	AAH98023	Murine 7-transmemb
	12	83.6	10.9	565	22	AAH07867	Human cDNA clone (
	13	83.6	10.9	1296	24	ABK49898	Human cDNA encodin
	14	83.6	10.9	1821	22	AAH15014	Human cDNA sequenc
	15	83.6	10.9	3610	24	ABN59933	Novel human coding
C	16	72.8	9.5	462	24	ABK53702	Human eosinophil-m
	17	64.4	8.4	2686	22	AAS27013	cDNA encoding nove
	18	64.4	8.4	3321	22	AAH44718	Human full-length
	19	64	8.3	1091	22	AAH27281	cDNA encoding nove
	20	60.8	7.9	1860	22	AAF99965	Human colon carcin
	21	60.8	7.9	5923	22	AAH02052	DNA encoding molec
C	22	60.8	7.9	5923	22	AAH02052	DNA encoding molec
	23	60	7.8	60	24	ABN41498	Human spliced tran
	24	56.2	7.3	365	14	AAQ39923	Expressed Sequence
	25	56.2	7.3	365	14	AAQ39935	Human brain expres
	26	53.2	6.9	417	22	ABA08384	Human collybistin
C	27	46.4	6.0	454	23	ABV54945	Human prostate exp
	28	46.4	6.0	541	24	ABN64838	Human cancer relat
	29	44.4	5.8	342	21	AAF08796	Fusarium venenatum
	30	44	5.7	1609	23	ABL25953	Drosophila melanog
	31	44	5.7	3381	23	ABL25950	Drosophila melanog
	32	44	5.7	3609	23	ABL25952	Drosophila melanog
	33	43.4	5.6	3786	23	AAH54228	Pseudomonas aerugi
C	34	42.4	5.5	1730	22	AAS27017	cDNA encoding nove
	35	42.4	5.5	1730	22	AAS27441	cDNA encoding nove
	36	42.4	5.5	1730	22	AAI62819	Human cDNA SEQ ID
C	37	42.4	5.5	1778	22	AAI60338	Human polynucleoti
	38	42.4	5.5	1816	22	AAI58552	Human polynucleoti
	39	42	5.5	3306	23	ABL12855	Drosophila melanog
	40	41.2	5.4	2159	15	AAQ67223	Mouse p55Nuc. Mus
	41	40.8	5.3	625	22	AAH07308	Human cDNA clone (
	42	40.8	5.3	1444	22	AAH41863	Human chondroadher
	43	40.8	5.3	1696	22	AAI93904	Human stomach canc
	44	40.8	5.3	1696	22	AAH16459	Human cDNA sequenc
	45	40.8	5.3	2172	23	AAH51456	Pseudomonas aerugi

ALIGNMENTS

RESULT 1

AAS64760

ID AAS64760 standard; cDNA; 3187 BP.

XX AAS64760;

AC AAS64760;

XX 13-FEB-2002 (first entry)

DT DNA encoding novel human diagnostic protein #564.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG00573.

XX New isolated polynucleotide and encoded polypeptides, useful in

Db 2314 AACTCAAGAAAGATTGATGGCAATTCACAACTTCTGTGTTCCGGGAAGGGATTCATCC 2373
QY 601 GTCTGGCGAGCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTCTGT 660
Db 2374 GTCTGGCGAGCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGT 2433
QY 661 TCAACGACGCTCTGCTATACAGAGCCGGGGGTGACGGCTCCCAATCAGTTTAAAGTCC 720
Db 2434 TCAACGACGCTCTGCTATACAGAGCCGGGGGTGACGGCTCCCAATCAGTTTAAAGTCC 2493
QY 721 ACGGCGAGCTCCGCTCTATGATGATGACGATTCAGGAGAGCGAAGACGA 769
Db 2494 ACGGCGAGCTCCGCTCTATGATGATGACGATTCAGGAGAGCGAAGACGA 2542

RESULT 4
ID AAC98992 standard; cDNA; 3094 BP.
XX AAC98992;
XX 09-MAR-2001 (first entry)
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX Homo sapiens.
OS
XX WO200055320-A1.
PN
XX 21-SEP-2000.
PD
XX 08-MAR-2000; 2000WO-US05989.
XX
PF
XX 12-MAR-1999; 99US-0124270.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PA
XX Rosen CA, Ruben SM;
PI
XX WPI; 2000-579444/54.
DR
XX P-PSDB; AAB54227.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 1; Page 664-665; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including

CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 3094 BP; 736 A; 849 C; 755 G; 749 T; 5 other;
Query Match 99.1%; Score 762; DB 21; Length 3094;
Best Local Similarity 100.0%; Pred. No. 1.5e-199; Mismatches 0; Gaps 0;
Matches 762; Conservative 0; Indels 0;
QY 8 GTGGTTTCAGAGCAGTGTGCAAAATTCATCTAATTTTCTCAAGAAATTCAT 67
Db 62 GTGGTTTCAGAGCAGTGTGCAAAATTCATCTAATTTTCTCAAGAAATTCAT 121
QY 68 ATTCCGGAATTTGAACCTTTGCAAAATTCATCTAATTTTCTCAAGAAATTCAT 127
Db 122 ATTCCGGAATTTGAACCTTTGCAAAATTCATCTAATTTTCTCAAGAAATTCAT 181
QY 128 ACGACTTGCCCTGTGGGAAGCGGCTCAATGCCCAATTCAGAGATTCACCAAGAAATCG 187
Db 182 ACGACTTGCCCTGTGGGAAGCGGCTCAATGCCCAATTCAGAGATTCACCAAGAAATCG 241
QY 188 CGATGTCTATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGGGCTCACTGTGAAGCA 247
Db 242 CGATGTCTATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGGGCTCACTGTGAAGCA 301
QY 248 CAGCAGGCGCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGGCGCTGGAGAACTT 307
Db 302 CAGCAGGCGCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGGCGCTGGAGAACTT 361
QY 308 CTGCAGAGACTTTGAGCTGCAGAAAGTGTTACCTACCCTCAACACCTTCTCTCTGG 367
Db 362 CTGCAGAGACTTTGAGCTGCAGAAAGTGTTACCTACCCTCAACACCTTCTCTCTGG 421
QY 368 GCCACTGCACCGGCTCATGCACATACAAGCAGGTCTCTGGAGCGGCTGTGCAAAACACACCC 427
Db 422 GCCACTGCACCGGCTCATGCACATACAAGCAGGTCTCTGGAGCGGCTGTGCAAAACACCC 481
QY 428 GCCGAGCCACGCCGACTTCAGGAGCTGCGGAGCCGCTTTGGCAGAGATCACGGAGATGT 487
Db 482 GCCGAGCCACGCCGACTTCAGGAGCTGCGGAGCCGCTTTGGCAGAGATCACGGAGATGT 541
QY 488 GGCACAGCTCCACGCTACGATGATCAAGATGGAGAAATTTCCAGAGCTCGCAGAACTCAA 547
Db 542 GGCACAGCTCCACGCTACGATGATCAAGATGGAGAAATTTCCAGAGCTCGCAGAACTCAA 601
QY 548 GAAAGATTTGATTGGCATTTGACAATCTTGTGTTCCGGGAAGGGAGTTTCATCCGTCTGG 607
Db 602 GAAAGATTTGATTGGCATTTGACAATCTTGTGTTCCGGGAAGGGAGTTTCATCCGTCTGG 661
QY 608 CAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACGCA 667
Db 662 CAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACGCA 721
QY 668 CGTCTGCTATACAGAGCGCGGGGCTGACGGCTCCATCATAGTTTAAAGTCCACGGGCA 727
Db 722 CGTCTGCTATACAGAGCGCGGGGCTGACGGCTCCATCATAGTTTAAAGTCCACGGGCA 781
QY 728 GCTCCGCTCTATGGCATGACGATTCAGGAGAGCGAAGACGA 769
Db 782 GCTCCGCTCTATGGCATGACGATTCAGGAGAGCGAAGACGA 823

RESULT 5
AAS58814/c
ID AAS58814 standard; cDNA; 426 BP.
XX AAS58814;
AC AAS58814;
XX 13-FEB-2002 (first entry)
XX

PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.	XX	
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.	XX	
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.	XX	
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.	XX	
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.	XX	
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.	XX	
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.	XX	
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.	XX	
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.	XX	
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250160.	XX	
PR	14-SEP-2000;	2000US-0232397.	PR	01-DEC-2000;	2000US-0250391.	XX	
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251030.	XX	
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0251988.	XX	
PR	14-SEP-2000;	2000US-0232400.	PR	05-DEC-2000;	2000US-0256719.	XX	
PR	14-SEP-2000;	2000US-0232401.	PR	06-DEC-2000;	2000US-0251479.	XX	
PR	14-SEP-2000;	2000US-0232403.	PR	08-DEC-2000;	2000US-0251856.	XX	
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251868.	XX	
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251869.	XX	
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.	XX	
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.	XX	
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.	XX	
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.	XX	
PR	25-SEP-2000;	2000US-0234998.	XX				
PR	26-SEP-2000;	2000US-0235484.	XX				
PR	27-SEP-2000;	2000US-0235834.	XX				
PR	27-SEP-2000;	2000US-0235836.	XX				
PR	29-SEP-2000;	2000US-0236327.	XX				
PR	29-SEP-2000;	2000US-0236367.	XX				
PR	29-SEP-2000;	2000US-0236368.	XX				
PR	29-SEP-2000;	2000US-0236369.	XX				
PR	29-SEP-2000;	2000US-0236370.	XX				
PR	02-OCT-2000;	2000US-0236802.	XX				
PR	02-OCT-2000;	2000US-0237037.	XX				
PR	02-OCT-2000;	2000US-0237038.	XX				
PR	02-OCT-2000;	2000US-0237039.	XX				
PR	02-OCT-2000;	2000US-0237040.	XX				
PR	13-OCT-2000;	2000US-0239935.	XX				
PR	13-OCT-2000;	2000US-0239937.	XX				
PR	20-OCT-2000;	2000US-0240960.	XX				
PR	20-OCT-2000;	2000US-0241221.	XX				
PR	20-OCT-2000;	2000US-0241785.	XX				
PR	20-OCT-2000;	2000US-0241786.	XX				
PR	20-OCT-2000;	2000US-0241787.	XX				
PR	20-OCT-2000;	2000US-0241808.	XX				
PR	20-OCT-2000;	2000US-0241809.	XX				
PR	20-OCT-2000;	2000US-0241826.	XX				
PR	01-NOV-2000;	2000US-0244617.	XX				
PR	08-NOV-2000;	2000US-0246474.	XX				
PR	08-NOV-2000;	2000US-0246475.	XX				
PR	08-NOV-2000;	2000US-0246476.	XX				
PR	08-NOV-2000;	2000US-0246477.	XX				
PR	08-NOV-2000;	2000US-0246478.	XX				
PR	08-NOV-2000;	2000US-0246523.	XX				
PR	08-NOV-2000;	2000US-0246524.	XX				
PR	08-NOV-2000;	2000US-0246525.	XX	</			


```
Db 133 ATGTTTCTGTTCTCAGATATGTTGCTGTACAAAGCAAAAGGAGTTGCAGGGACCAAGC 192
Qy 708 CAGTTTAAAGTCCAGCGCAGCTCCCGCTCTATGCGATGACGATTGAGGAGCAAGAC 767
Db 193 CACTTCCGGATCCGGGGCTCCTTCCCTCCAAAGCATGCTGGTGAAGAAAGTGATTAAC 252
Qy 768 GA 769
Db 253 GA 254

RESULT 7
AAS27440
ID AAS27440 standard; cDNA; 716 BP.
XX AC
AAS27440;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seq ID 475.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-MAR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
```


CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 717 BP; 170 A; 196 C; 203 G; 144 T; 4 other;

Query Match 12.7%; Score 98; DB 24; Length 717;
Best Local Similarity 65.6%; Pred No. 6.7e-17;
Matches 143; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 552 GATTGATTGGCATTGACAAATCTTGTTCCGGAAGGAGTTTCATCGTCTGGGCAGC 611
II III III III III III III III III III III III III III III III
Db 28 GACCTGGTGGCATAGAGAACCTCTGCTCTGCGCAGGAGTTTCATCGTCTGAGGCTGC 87

QY 612 CTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGAGCTC 671
II III III III III III III III III III III III III III III III
Db 88 CTTCAACAAGCTCACCAGAAGGGCTGCGCAGCAGAGGATGTTTTTTCTGTTTCAGATATG 147

QY 672 CTGCTATACACAGCGGGGGCTGACGGGCTCCAAATCACTTTTAAAGTCCACGGGAGCTC 731
II III III III III III III III III III III III III III III III
Db 148 TTGCTGTACAGAGCAAGAGATTGCGAGGACCGCAGCTTCGGATCCGGGGCTCTCTT 207

QY 732 CGGCTCTATGGCATGACGATTGAGGAGCGCAAGACGA 769
II III III III III III III III III III III III III III III III
Db 208 CCCCTCCAAGGCATGCTGTGGGAAGAAAGTGATAACGA 245

RESULT 9
ABLO3867
ID ABL03867 standard; cDNA; 1747 BP.
XX
AC ABL03867;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 6083.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB59764.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 6083; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA

CC sequences (ABLO1840-ABLI6175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1747 BP; 431 A; 492 C; 461 G; 363 T; 0 other;

Query Match 12.6%; Score 97.2; DB 23; Length 1747;
Best Local Similarity 50.8%; Pred No. 1.7e-16;
Matches 259; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 171 GATTACCAAGAATCGCGCATCTCATGTGAAGAACATTTACGGGCATGAAGCACCTGGCG 230
II III III III III III III III III III III III III III III III
Db 516 GAAGCCCAACCGCATCGGAGACGTCATGATGAAGCACATGGCGCTCTGCCCATCTACGAT 575

QY 231 GCTCACCCTGTGAACACACAGCAGCCCTTGGAGGCCCTTGGAGAAATGGAATCAAGAGCTCC 290
II III III III III III III III III III III III III III III III
Db 576 GAGTACGTGCAGACGACCTGGACATCTTGCACCTGTATGAAGGACATGTACGAAGCGCAT 635

QY 291 CGCGCGCTGGAGAACTTCTGCAGAGACTTTTGAGCTGCAGAGGTGTGTACCTACCGGTC 350
II III III III III III III III III III III III III III III III
Db 636 GAACGCTTCCGTCAGGTGTACAGAGGATTTGAGCAGCAAAAGGTTTGTCTATCACCATC 695

QY 351 AACACCTTCTCTCGCGCCACTGCACCGGCTCATGCATACAAGCAGTCTCTGGAGCGG 410
II III III III III III III III III III III III III III III III
Db 696 GCGGAACCTTCTACTGAAGCCCTCAACGCCCTGTTGCACTACCAACTGATCTTAGAGCGG 755

QY 411 CTGTCAAAACACCACCCGCCGAGCCAGCCGACTTCAGGGGACTGCGGAGCCCTTTGGCA 470
II III III III III III III III III III III III III III III III
Db 756 CTCTCGCACTACTATGGGAGGAGCATATCGACTATGCGGATGCTATGGCCGTGCACCAC 815

QY 471 GAGATCAGGAGATGGTGGCACAGCTCCACGATCAGATGATCAAGATGGAGAAATTTCCAG 530
II III III III III III III III III III III III III III III III
Db 816 TTGCTCGTTCGCGACCAACCAAGGTTATTAGTCCCGACTTCCCGACTCTGCACAACTTTGTG 875

QY 531 AAGCTGCAGAACTCAAGAAAGATTTGATTTGGCATTTGACAATCTTTGTGTTCCGGGAAGG 590
II III III III III III III III III III III III III III III III
Db 876 GAGCTGTGTGAAT---GCAACGCGACATCAACTTCGAGCAGTTGGTTCAGCCCCCATCGC 932

QY 591 GAGTTTCATCCGCTCTGGCAGCCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATG 650
II III III III III III III III III III III III III III III III
Db 933 CGCCTCATCCGCGAGGATGCCCTTCTGAAGCACTCCAAAGCGCGGTCTGCAGCAGAGGATG 992

QY 651 TTCTTCTCTTCAACGAGCTCTCTGCTATAC 680
II III III III III III III III III III III III III III III III
Db 993 TTCTTCTTGTCTCCGACCTGCTCTCTAC 1022

RESULT 10
ABLO3866/C
ID ABL03866 standard; cDNA; 4150 BP.
XX
AC ABL03866;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 6080.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX


```
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
OS EP1074617-A2.
PN 07-FEB-2001.
PD 28-JUL-2000; 2000EP-0116126.
PF 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 1; SEQ ID 4702; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides, and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX Sequence 665 BP; 209 A; 138 C; 154 G; 161 T; 3 other;
SQ Query Match 10.9%; Score 83.6; DB 22; Length 665;
Best Local Similarity 51.3%; Pred. No. 6e-13;
Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
QY 338 TTACCTACCGCTCAACACCTTCTCTCTGGCCACTGACCGGCTCATGACATACAGCA 397
Db 27 TAATCTGGCCCTCAAGCACTACTGTCTCAAGCGGTTACAGAGATCCCCACGTACAGGCT 86
QY 398 GGTCTCTGAGCGGCTGTGCAACACCAACCCGCCGAGCCGACCTTCAGGAGCTGCCG 457
Db 87 GTTGCTGACAGATATTTCAGAAATCTCATAGAAGATGCTGGAGATTACAGACACTCA 146
QY 458 AGCCGCTTTGGCAGAGATCAGGAGATGTGGCAGACAGCTCCACGGTACGATGATCAAGAT 517
Db 147 AGATGCCCTTGCTGTTGTATAGAGTAGTCCCAACCAACGCGCAATGACACCATGAAGCAAG 206
QY 518 GGAGAAATTCAGAGACTCCGAACTCAAGAAAGATTTGATGGCATTTGACAATCTTGT 577
```

```
Db 207 AGACAACTTTTCAGAAACTTATGCAAAATTCAGTACAGCTTAAATGGACACCATGAAATTGT 266
QY 578 GGTTCGGGAAGGAGTTCATCGTCTGGCGACGCTCAGCAAGCTCTCGGGAAGGGCT 637
Db 267 GCAGCTGTGTCGGGTTTCTCAAGNAGGAATTCGTGATGAGCTCTCTCGGAAGTGAT 326
QY 638 CCAGCAGCGCATGTTCTCTCTTCAACGACGCTCTGCTATACACGAGCGGGGGCTGAC 697
Db 327 GCAACCTCGAATGTTTCTCTGTTTAATGATGCCCTGCTGTATACAAACACCATGCGAGTC 386
QY 698 GGCCTCCAATCAGTTTAA 715
Db 387 TGGGATGTATAAACTGAA 404
RESULT 13
ABK49898
ID ABK49898 standard; cDNA; 1296 BP.
XX
AC ABK49898;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human cDNA encoding Faciogenital dysplasia 1-like protein.
XX
KW Human; ss; gene; faciogenital dysplasia; FGDL-like protein;
KW Aarskog syndrome; X-linked developmental disorder; cancer; obesity;
KW guanine nucleotide exchange factor; overweight; anorexia; cachexia;
KW wasting disorder; appetite suppression; appetite enhancement; satiety;
KW modulation of body weight; eating disorder; bulimia; hypertension;
KW type 2 diabetes; coronary artery disease; hyperlipidaemia;
KW stroke; gallbladder disease; gout; osteoarthritis; sleep apnea;
KW respiratory problem; thrombotic disease; polycystic ovarian syndrome;
KW reduced fertility; complication of pregnancy; menstrual irregularity;
KW hirsutism; stress incontinence; depression.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..1296
FT /*tag= a
FT /product= "FGDL-like protein"
FT /partial
FT /note= "No stop codon shown"
PN WO200224731-A2.
PD 28-MAR-2002.
PF 24-SEP-2001; 2001WO-EP11009.
PR 25-SEP-2000; 2000US-234983P.
XX (FARB ) BAYER AG.
XX
XX Kossida S;
XX WPI: 2002-383181/41.
XX P-PSDB; AAU80165.
XX
XX New human faciogenital dysplasia (FGDL)-like proteins, regulators of
XX which are useful for preventing and treating cancer, obesity, cachexia,
XX depression, diabetes, hypertension and stroke -
XX Claim 1; Fig 1; 78pp; English.
XX
XX The invention relates to a purified human faciogenital dysplasia (FGDL)-
XX like protein and its encoding cDNA. Faciogenital dysplasia (also known as
XX Aarskog syndrome) is an X-linked developmental disorder and is a guanine
XX nucleotide exchange factor. Also included are a hybridisable sequence,
XX fragment, derivative, allelic variant, or a sequence which deviates
XX from the cDNA (or sequence 50% identical to it) due to degeneration of
```

CC the genetic code, an expression vector comprising the nucleotide
 CC sequence, a host cell containing the vector, modulators of the activity
 CC of the protein, fusion proteins with the FGDI-like protein, and screening
 CC for agents which modulate/reduce the activity of the protein.
 CC The FGDI-like protein and cDNA are useful for screening for agents which
 CC decrease the activity of FGDI-like protein. The cDNA is useful for
 CC detecting a polynucleotide encoding a FGDI-like protein in a biological
 CC sample. The protein modulators and isolated agents are useful for
 CC treating a FGDI-like protein dysfunction related disease, such as cancer,
 CC obesity, overweight, anorexia, cachexia, wasting disorders, appetite
 CC suppression, appetite enhancement, increases or decreases in satiety,
 CC modulation of body weight, and/or other eating disorders such as
 CC bulimia, obesity/overweight-associated co-morbidities including
 CC hypertension, type 2 diabetes, coronary artery disease, hyperlipidaemia,
 CC stroke, gallbladder disease, gout, osteoarthritis, sleep apnea and
 CC respiratory problems, endometrial, breast, prostate, colon cancer,
 CC thrombotic disease, polycystic ovarian syndrome, reduced fertility,
 CC complications of pregnancy, menstrual irregularities, hirsutism, stress
 CC incontinence and depression. The coding sequence of FGDI-like protein
 CC polynucleotide is useful for generating antisense oligonucleotides or
 CC ribozymes. These antisense oligonucleotides are useful for modulating
 CC FGDI-like protein gene expression. The FGDI-like protein is useful for
 CC generating antibodies against FGDI-like protein amino acid sequences and
 CC for use in various assay systems. The present sequence is the cDNA
 CC encoding the human FGDI-like protein.
 XX
 SQ Sequence 1296 BP; 415 A; 273 C; 281 G; 327 T; 0 other;

Query Match 10.9%; Score 83.6; DB 24; Length 1296;
 Best Local Similarity 51.3%; Pred. No. 8.1e-13;
 Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
 QY 338 TTACCTACCGCTCAACACCTTCTCTCGGCGCACTGCACCGGCTCATCTACTACAAGCA 397
 Db 18 TAACTGGCCCTCAGCACTACTCTGCTCAAGCGGTTTCAGAGATCCCCAGTACAGGCT 77
 QY 398 GGTCTCGGAGCGGCTGTGCAACACACACCGCCGAGCCCGACTTCAGGAGTGGCG 457
 Db 78 GTTGTGACAGATTAATTTGAAGAATCTCATAGAAGATGCTGGAGATTACAGAGACACTCA 137
 QY 458 ACCCGCTTTGGCAGAGATCAGGAGATGCTGGCAGAGTCCACGCTCCAGTATCAAGAT 517
 Db 138 AGATGCCCTTGGCTTTGTTATAGAGTAGTCCCAACCGCCCAATGACACCATGAAGCAAGG 197
 QY 518 GGAGATTTTCCAGAGCTCAGCAAGTCAAGAAAGATTTGATGGATTCGAACTTCTTGT 577
 Db 198 AGACAACTTTCAGAACTTATGCAATTCAGTACAGCTTAATGGACACCATGAATTTGT 257
 QY 578 GGTTCGGGAAGGAGTTCATCCGCTGGCGAGCTCAGCAAGCTCTCGGGGAAGGGGCT 637
 Db 258 GCAGCCTGCTCGGGTTTTTCTCAAAGAAGGAATTTCTGATGAAGCTCTCTCGGAAAGTGAT 317
 QY 638 CCAGCAGCGCATGCTTCTCTTCAACGACGTCCTGCTATATACAGCGCGGGGCTGAC 697
 Db 318 GCAACCTCAAAATGTTTCTCTGTTAATGATGCCCTGCTGTATACAAACACCATGCACTC 377
 QY 698 GGCCTCCAACTAGTTAA 715
 Db 378 TGGATGTATAACTGAA 395

RESULT 14
 ID AAH15014
 XX AAH15014 standard; cDNA; 1821 BP.
 AC AAH15014;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:12971.
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX

OS Homo sapiens.
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-01161126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 12971; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1821 BP; 594 A; 344 C; 369 G; 514 T; 0 other;

Query Match 10.9%; Score 83.6; DB 22; Length 1821;
 Best Local Similarity 51.3%; Pred. No. 9.4e-13;
 Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
 QY 338 TTACCTACCGCTCAACACCTTCTCTCGGCGCACTGCACCGGCTCATCTACTACAAGCA 397
 Db 27 TAACTGGCCCTCAAGCACTACTCTGCTCAAGCGGTTTCAGAGATCCCCAGTACAGGCT 86
 QY 398 GGTCTCGGAGCGCTGTGCAACACACCGCGCGAGCCAGCTTCAGGAGCTGCGG 457
 Db 87 GTTGTGACAGATTATTTGAGAATCTCATAGAAGATGCTGGAGATTACAGAGACACTCA 146
 QY 458 AGCCGCTTTGGCAGAGATCAGGAGATGGTGCAACAGCTCCACAGGTACGATGATCAAGAT 517
 Db 147 AGATGCCCTTGTCTGTTGTTATAGAGTAGTGAACCAACCGCAATGACACCATGAAGCAAG 206
 QY 518 GGAGATTTCCAGAGCTGCAGGAATCAAGAAAGATTTGATGGCATTGCAATCTTGT 577
 Db 207 AGACAACCTTCAGAAAACCTTATGCAAAATTCAGTACAGCTTAAATGACACCATGAAATTTGT 266

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 04:45:14 ; Search time 72 Seconds
(without alignments)
3275.478 Million cell updates/sec

Title: US-09-555-342b-1_COPY_1733_2501

Perfect score: 769
Sequence: 1 tcaactcgtggttcagagc.....attgaggagagcgagagca 769

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2.6/ptodata/2/ina/PCUS_COMB.seq: *
6: /cgn2.6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.8	5.3	4776	US-08-852-401-1	Sequence 1, Appli
2	39.6	5.1	4722	US-08-979-608A-14	Sequence 14, Appli
3	37.2	4.8	44377	US-08-804-227C-7	Sequence 7, Appli
4	37.2	4.8	44377	US-08-804-198-1	Sequence 1, Appli
5	36.4	4.7	2824	US-09-010-928B-3	Sequence 3, Appli
6	36	4.7	2214	US-08-864-038A-1	Sequence 1, Appli
7	36	4.7	3331	US-08-864-038A-2	Sequence 2, Appli
8	36	4.7	3331	US-08-864-038A-4	Sequence 4, Appli
9	35.4	4.6	2132	US-09-552-322-1	Sequence 1, Appli
10	35.4	4.6	4403765	US-09-103-840A-2	Sequence 2, Appli
11	35.4	4.6	4411529	US-09-103-840A-1	Sequence 1, Appli
12	34.6	4.5	657	US-09-527-345-3	Sequence 3, Appli
13	34.6	4.5	1029	US-08-911-853-1	Sequence 1, Appli
14	34.6	4.5	1029	US-09-479-409-1	Sequence 1, Appli
15	34.6	4.5	1029	US-09-479-453-1	Sequence 1, Appli
16	34.6	4.5	4377	US-08-911-853-28	Sequence 28, Appli
17	34.6	4.5	4377	US-09-479-409-28	Sequence 28, Appli
18	34.6	4.5	4377	US-09-479-453-28	Sequence 28, Appli
19	34.6	4.5	10348	US-08-457-273B-41	Sequence 41, Appli
20	34.6	4.5	10348	US-08-556-419-13	Sequence 13, Appli
21	34.6	4.5	10348	US-09-041-886-14	Sequence 14, Appli
22	34.6	4.5	10366	US-08-246-982A-5	Sequence 5, Appli
23	34.6	4.5	10366	US-08-453-265-5	Sequence 5, Appli
24	34.4	4.5	7042	US-09-092-508-1	Sequence 1, Appli
25	34.4	4.5	7042	US-09-435-115-1	Sequence 1, Appli
26	34.4	4.5	7042	US-09-098-310-1	Sequence 1, Appli
27	34.4	4.5	7042	US-09-690-364-21	Sequence 21, Appli

c 28 34.4 4.5 7075 4 US-09-092-508-15 Sequence 15, Appli
c 29 34.4 4.5 7075 4 US-09-435-115-15 Sequence 15, Appli
c 30 34.2 4.4 1331 4 US-09-335-983-14 Sequence 14, Appli
31 34.2 4.4 4195 1 US-08-340-011-1 Sequence 1, Appli
32 34.2 4.4 4195 3 US-08-901-710-1 Sequence 1, Appli
33 34.2 4.4 4416 3 US-08-795-430-1 Sequence 1, Appli
34 34.2 4.4 4416 4 US-09-355-700-1 Sequence 36, Appli
35 34.2 4.4 4416 4 US-08-601-132-36 Sequence 31, Appli
36 34.2 4.4 4425 1 US-08-222-616-31 Sequence 31, Appli
37 34.2 4.4 4425 4 US-08-446-648-31 Sequence 31, Appli
38 34.2 4.4 4425 5 PCT-US95-04228-31 Sequence 31, Appli
39 34.2 4.4 4795 1 US-08-340-011-3 Sequence 3, Appli
40 34.2 4.4 4795 3 US-08-901-710-3 Sequence 3, Appli
c 41 34.2 4.4 5349 4 US-09-068-101-7 Sequence 7, Appli
42 34.2 4.4 5348 3 US-08-894-440-1 Sequence 1, Appli
43 34.2 4.4 5348 3 US-08-817-188-2 Sequence 2, Appli
44 34.2 4.4 5348 4 US-09-458-093-1 Sequence 1, Appli
45 34.2 4.4 9108 4 US-08-446-648-45 Sequence 45, Appli

ALIGNMENTS

RESULT 1
US-08-852-401-1
; Sequence 1, Application US/08852401
; Patent No. 5976836
; GENERAL INFORMATION:
; APPLICANT: Weber, J. Mark
; APPLICANT: Hessler, Paul E.
; APPLICANT: Larsen, Peter E.
; APPLICANT: Luu, B. Minh
; TITLE OF INVENTION: Methods and Compositions for Enhancing
; TITLE OF INVENTION: Erythromycin Production
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.
; STREET: 2 Prudential Plaza, 180 N. Stetson, Suite
; STREET: 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,401
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueller, Lisa L.
; REGISTRATION NUMBER: 38,978
; REFERENCE/DOCKET NUMBER: FER2159P00300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5400
; TELEFAX: 312-616-5460
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-852-401-1

Query Match 5.3%; Score 40.8; DB 2; Length 4776;
Best Local Similarity 43.9%; Pred. No. 0.058;
Matches 174; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 280 TCAAGAGCTCCCGCGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTTGTT 339

Db 797 TCACCGCGCGGGCGGCATCAAGAGGAGACACCGCGAGCGCGTGTGGCGCGTCTCAACG 856
QY 340 ACTACCGCTCAACACTTCTCTCCCTGCGCGCACTGACCGCGCTCATGCACTACAAGCAGG 399
Db 857 AGATCGCGCACACCGCGCGCATAGGTGCGCTGCGCGCACCGCGGGTAAATCGGCGCTGC 916
QY 400 TCTGAGCGCGGTGTGCAACACACCGCGCGCGCGCGCGCGCTTCAGGGGACTGCGGAG 459
Db 917 TGGTGGCGGAGCTGCCACACCGCGGTGTTCCCGCGCTTCGCCGAGCGCTTGGAGGCGCGG 976
QY 460 CGCTTTGGCAGAGATACGAGAGATGTTGGCAGACGCTCCACCGTACGATGATCAAGATGG 519
Db 977 CGTGGCGCGCGGTACGCTCGCTCTGCAACACCGCGCGCTGCGGATGACGAGGAGG 1036
QY 520 AGAATTTCCAGAGCTGCAGCACTCAAGAAAGATTTGATTTGGCATTCATCTTGGG 579
Db 1037 ACTAGCTCGGATGCTCATCGCGCGCGCGGTGAGGCGATGCTTCTGCTGCGCGGAGA 1096
QY 580 TTCCGGGAAGGAGTTTCATCCGCTCTGGGCGAGCTCAGCAAGCTCTCGGGGAAGGGCTCC 639
Db 1097 TGCCCAACACCGCGCGGAGCGGATCAGCGCGAGCTACTACGAGAAGCTGCTGGCGG 1156
QY 640 AGCAGCGCATGTTCTTCCTGTTTCAAGCAGCTCTCTCG 675
Db 1157 ACGGCGTGGCGCATGCTTCTCGTCAACGCGCGCGCGC 1192

RESULT 2

US-08-979-608A-14
; Sequence 14, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59810)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 61...1731
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-08-979-608A-14
Query Match 5.1%; Score 39.6; DB 4; Length 4722;
Best Local Similarity 47.5%; Pred. No. 0.13;
Matches 154; Conservative 0; Mismatches 164; Indels 6; Gaps 1;
QY 218 GAAGCACTGGCGGCTCACCTGTGGAAGCACACGAGGCGCTTTGGAGCGCCTGGAGAATGG 277
Db 618 GGAGAAGCTGGCGGCTGTGTGCAAGAAGTATGCGGAACCTGCTCGAGGAGACCGGAATC 677
QY 278 ATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAAGACATTTGAGCTGCAAGAGGTGG 337
Db 678 GCAGAAGCAGATGAAGCTGCTGCAAGAAGCAGACCGAGCTGGTGCAGGAGAGG---- 733
QY 338 TTACTACCGCTCAACACTTCTCTCTCGGCGCACTGCACCGCTCATGCACTACAAGCA 397
Db 734 --ACACCTGCTGGCGAGCACAGCAAGCCATCTCTGCCCCCAGCAAGCTCGAGAGCT 791
QY 398 GGTCTTGGAGCGGCTGTGCAACACACCGCGCGAGCCACCGCTTTCAGGGACTGCCG 457
Db 792 GTGCGCGGAGCTGACGCGCACACCGCTCGCTCAAGGAAGAAGCTGTGCACGAGCGCG 851
QY 458 AGCCGCTTTGGCAGAGATCAGGGAGATGTTGGCAGCAGCTCCACGCTAGCATGATCAAGAT 517
Db 852 AGAGGAGGAGAGAGCGCAAGGAGGTGACGTCAACATTTCCAGATGACGCTCAACGACAT 911
QY 518 GGAGAATTTCCAGAAGCTGCACGA 541
Db 912 TCAGCTGCAGATGGAGCAGCACAA 935

RESULT 3

US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

Query Match          4.8%; Score 37.2; DB 2; Length 44377;
Best Local Similarity 43.8%; Pred. No. 2;
Matches 162; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 147 GGCCTCAAAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGTCATGCTGAAGAAC 206
Db 29515 GACTTCGGACCGCCTTACCGGAGCGGACGCTCCCGCGGACGCTGCTCTGAAC 29574

QY 207 ATTCAGGGCATGAAGACCTCGCGGCTACCTGTGGAACACAGCGAGCCCTTGAGGCC 266
Db 29575 TCGCTCACCAGGAGTTCGTGGACGCTCCCTCGGCTGCTCCGCGCGCGGTTTC 29634

QY 267 CTGGAGATGGAATCAAGAGCTCCCGGGCTGGAGAACTTCTGCAGAGACTTTGAGCTG 326
Db 29635 CTGGAGCTGGCAAGACCGAGCTCCGGGACCCCGGAGCGGATCGCCGCCAACAACCCCGG 29694

QY 327 CAGAAGGTGTGTACTACCTACCGCTCAACACCTTCCTCTCGCGCCACTGCACCGGCTCATG 386
Db 29695 GTGCGCTACCGGCGTTCACCTCAACGAGCGCGGACCGACACTCGCGCGGCTGCTG 29754

QY 387 CACTACAAGCAGTCTTGAGCGGTGTGCAACACCAACCGCGGAGCGACGCGACTTC 446
Db 29755 CGGGAACATGATGACCTGTTCGCGCGCGGCTGCTGCACCCGCTGCGCTGCACCCAC 29814

QY 447 AGGAGCTCGCGAGCGCTTTGGCAGAGATCAGGAGATGGTGCAGAGCTTCACAGGTAGC 506
Db 29815 GACGTGCGCGCGCGGCGGACGCTTCGCGCACCATCAGCCAGCGCGCGGACACCGGAAG 29874

QY 507 ATGATCAAGA 516
Db 29875 CTCGTCCTGA 29884
```

RESULT 4

```
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Bargett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostek, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN USA
; COUNTRY: IN
; ZIP: 46285
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1
```

Query Match 4.8%; Score 37.2; DB 2; Length 44377;

Best Local Similarity 43.8%; Pred. No. 2;

Matches 162; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```
QY 147 GGCCTCAAAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGTCATGCTGAAGAAC 206
Db 29515 GACTTCGGACCGCCTTACCGGAGCGGACGCTCCCGCGGACGCTGCTCTGAAC 29574

QY 207 ATTCAGGGCATGAAGACCTCGCGGCTACCTGTGGAACACAGCGAGCCCTTGAGGCC 266
Db 29575 TCGCTCACCAGGAGTTCGTGGACGCTCCCTCGGCTGCTCCGCGCGCGGTTTC 29634

QY 267 CTGGAGATGGAATCAAGAGCTCCCGGGCTGGAGAACTTCTGCAGAGACTTTGAGCTG 326
Db 29635 CTGGAGCTGGCAAGACCGAGCTCCGGGACCCCGGAGCGGATCGCCGCCAACAACCCCGG 29694

QY 327 CAGAAGGTGTGTACTACCTACCGCTCAACACCTTCCTCTCGCGCCACTGCACCGGCTCATG 386
Db 29695 GTGCGCTACCGGCGTTCACCTCAACGAGCGCGGACCGACACTCGCGCGGCTGCTG 29754

QY 387 CACTACAAGCAGTCTTGAGCGGTGTGCAACACCAACCGCGGAGCGACGCGACTTC 446
Db 29755 CGGGAACATGATGACCTGTTCGCGCGCGGCTGCTGCACCCGCTGCGCTGCACCCAC 29814

QY 447 AGGAGCTCGCGAGCGCTTTGGCAGAGATCAGGAGATGGTGCAGAGCTTCACAGGTAGC 506
Db 29815 GACGTGCGCGCGCGGCGGACGCTTCGCGCACCATCAGCCAGCGCGCGGACACCGGAAG 29874

QY 507 ATGATCAAGA 516
Db 29875 CTCGTCCTGA 29884
```

RESULT 5

US-09-010-928B-3/c

```
; Sequence 3, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: 1..2824
; LOCATION: 1..2824
; OTHER INFORMATION: /note= "Flagelliform DNA sequence
; OTHER INFORMATION: taken from 3' region. Stop codon begins at position 2722."
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2724
; US-09-010-928B-3

Query Match 4.7%; Score 36.4; DB 2; Length 2824;
Best Local Similarity 48.5%; Pred. No. 0.84;
Matches 100; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 340 ACTACCGCTCAACACCTTCTCTCGCGCCACTGCACCGGCTCATGCACTACAGCAGG 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 678 ACTCTGCTCGCGCTGCTCTCTCGCGCCACTGCACCGGCTCATGCACTACAGCAGG 619
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 400 TCGTGGAGCGGTGTGCAACACCCCGCGGAGCCAGCCAGCTTCAGGGAGCTGCCGAG 459
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 618 TCCAGCACCAACAGGTCAGCTCCACAGGTCAGCAGCACCAGGTCAGCTCCACAGG 559
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 460 CCGCTTGGCAGATCAGGAGATGGTGGCAGAGCTCCACAGGTCAGGATGATCAAGATGG 519
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 558 TCCAGCACCAACAGGTCAGGTCAGCAGCACCAGCAGCAGGTCAGCAGCAGGTCAGCAGG 499
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 520 AGAATTTCCAGAGCTGCACGAATC 545
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 498 TCCAGCACCAACAGGTCACCAACAC 473
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-08-864-038A-1/c
; Sequence 1, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
```

```
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2214
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; US-08-864-038A-1

Query Match 4.7%; Score 36; DB 3; Length 2214;
Best Local Similarity 55.6%; Pred. No. 0.97;
Matches 69; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 323 GCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCTCTCGCGCCACTGCACCGGCT 382
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1073 GCTGCAGCGGCGAGCGGCGAGCAGCGCTCGCGCTCTCTCTCCACCACTCCAGCTCT 1014
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 383 CATGCACCTACAGCAGGTCTGTGAGCGGCTGTGCAACACCAACCGCGCGGAGCGCCGA 442
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1013 CGGCCACCAACCGGCTGCGCGGCGAGCTGCAGCGGCGAGCAGCAGCTGTGCGG 954
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 443 CTTC 446
|||||
Db 953 CTC 950

RESULT 7
US-08-864-038A-2/c
; Sequence 2, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
```

IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-2

	Query Match	4.7%	Score 36;	DB 3;	Length 3331;
	Best Local Similarity	55.6%;	Pred. No. 1.2;		
	Matches 69;	Conservative	0;	Mismatches 55;	Indels 0;
				Gaps 0;	
QY	323	GCTCGAAGGTGTACCTACCGCTCAACACCTTCTCTCGTGGCCACATGCACCGGCT	382		
Db	1122	GCTCGAGGGCAGCGGAGCAGCAGCGCTCCGCGCTCTCTCCACCACTCCACCTCT	1063		
QY	383	CATCGACTACAAGCAGGTCTCTGGACGGCTGTGCAAAACACCAACCGCGGACGCGCGA	442		
Db	1062	CCGCCACCAACCGCGGTGCGCGCGCAGCTGCACGAGCGGCGAGCAGCAGCTCTGCGG	1003		
QY	443	CTTC	446		
Db	1002	CCTC	999		

```

RESULT 9
US-09-552-322-1
; Sequence 1, Application US/09552322
; Patent No. 6436642
; GENERAL INFORMATION:
; APPLICANT: Gould-Rothberg
; APPLICANT: Rastelli
; TITLE OF INVENTION: METHOD OF CLASSIFYING A THYROID CARCINOMA USING
; FILE OF INVENTION: DIFFERENTIAL GENE EXPRESSION
; FILE REFERENCE: 15966-548
; CURRENT APPLICATION NUMBER: US/09/552.322
; CURRENT FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/130,123
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: 60/193,203
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2132
; TYPE: DNA

```

```
; ORGANISM: Homo sapiens
US-09-552-322-1

Query Match          4.6%; Score 35.4; DB 4; Length 2132;
Best Local Similarity 55.2%; Pred. No. 1.4;
Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 607 GCAGGCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCACAG 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 GGGGGCGGCTGAGCTCTCGGCTCGGCTTCGACGAGCGGCAACTCTCGCTGCTCATCC 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 667 AGCTCTCTGTATACAGCAGCGCGGGGCTGACGGCTTCAATCAGTTTAAAGTCCACGGGC 726
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 GCGCGGTGGAGGAGAGCGGAGCGGGGCTGTACACCTGCAACCTGCACCATCACTACTGCC 295
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 727 AGCTC 731
    ||| |||
Db 296 ACCTC 300

RESULT 10
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          4.6%; Score 35.4; DB 4; Length 4403765;
Best Local Similarity 52.3%; Pred. No. 43;
Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 345 CCGCTCAACACCTTCTCTCGGGCACTGCACCGGCTCATGCAAGCAGGTCCTG 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391026 CTGCGGTAGACGGTTATCCGGGTATGCGAGTAGTTGCACATGATCGACACCCAGGACCTG 391085

QY 405 GAGCGGCTGTGCAAAACACACCCCGGCGGACCTTCAGGGACTTCAGGACCGCT 464
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391086 TCCCGCAGGCGGAGTCTCTCCCGGCGACCTCACCGAGGTTTCGGTCCGCGCCGAG 391145

QY 465 TTGGCAGAGATCAGGAGATGTGGCACA 493
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391146 TGTGGAGGCTCGCGGTGGGTGAAGCACA 391174

RESULT 11
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
```

```
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          4.6%; Score 35.4; DB 4; Length 4411529;
Best Local Similarity 52.3%; Pred. No. 43;
Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 345 CCGCTCAACACCTTCTCTCGGGCACTGCACCGGCTCATGCAAGCAGGTCCTG 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390971 CTGCGGTAGACGGTTATCCGGGTATGCGAGTAGTTGCACATGATCGACACCCAGGACCTG 391030

QY 405 GAGCGGCTGTGCAAAACACACCCCGGCGGACCTTCAGGGACTTCAGGACCGCT 464
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391031 TCCCGCAGGCGGAGTCTCTCCCGGCGACCTCACCGAGGTTTCGGTCCGTCGCGCCGAG 391090

QY 465 TTGGCAGAGATCAGGAGATGTGGCACA 493
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391091 TGTGGAGGCTCGCGGTGGGTGAAGCACA 391119

RESULT 12
US-09-527-345-3
; Sequence 3, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
; CURRENT FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc_feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-527-345-3

Query Match          4.5%; Score 34.6; DB 4; Length 657;
Best Local Similarity 36.5%; Pred. No. 1.3;
Matches 61; Conservative 9; Mismatches 97; Indels 0; Gaps 0;

QY 218 GAAGCACCTGCGCGCTCACCTGTGGAAGCACAGCGAGCGCTTGGAGCGCTGGAGATGG 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 439 GARCNGCNGCNGCNGCNGCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGTNGNGCN 498

QY 278 AATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAAGACTTTTGAGTCGACAGAGGTGTG 337
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 499 GARCNGCNGCNGARCCNGCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGTNGNGNTN 558

QY 338 TTACCTACCGCTCAACACCTTCTCTCGGCGCACTGCACCGGCTCA 384
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 559 GARCNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNA 605
```



```

; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,409
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-479-409-1

Query Match 4.5%; Score 34.6; DB 4; Length 1029;
Best Local Similarity 46.2%; Pred. No. 1.7;
Matches 115; Conservative 0; Mismatches 134; Indels 0; Gaps

Qy 271 AGAATGGAATCAAGAGCTCCGGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGA 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29 AGGAAGTGCTGATGTGAAACCGCCCATGGAGGAACCTACCCGGCATCAGGCGCGCAGCAGG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 331 AGTGCTGTTACCTACCGCTCAACACCTTCTCTCTGGGGCCACTGCACCGGCTCATGCACT 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 TGTGCGGCTCGCGCTGCTCAGGCTGGAGACCCCTCGCGCGAGCTGCTGACGAGCTTCA 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 391 ACAAGCAGGTCCTGGAGCGCTGTGCAACACACCCGCGGAGCCAGCCGACTTCAGG 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 TCGCCCAAGAGGAGGACCTGCACAAGCAGCACTGCAACTGGACGGGCGGAGTGCCT 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 451 ACTGCCAGGCGCTTTGTCAGAGATCAGGAGATGTTGGCAGACAGCTCCACGGTACGATGA 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 GGCTCACTGTCACAGGCGGCCATCGACCAACCGCTGGCGCGGCGGAGCGGCTG 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 511 TCAAGATGG 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 TGCTGCTGG 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-09-479-453-1
; Sequence 1, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37

```


GenCore version 5.1.3
Cop'right (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 05:01:45 ; Search time 97 Seconds
(without alignments)
3135.150 Million cell updates/sec

Title: US-09-555-342B-1_COPY_1733_2501

Perfect score: 769

Sequence: 1 tcaacttggtggttcagagc.....attgaggagagcgagacga 769

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 355320 seqs, 197730502 residues

Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	762	99.1	3094	10	US-09-925-297-220
2	248.6	32.3	408	10	US-09-783-590-3620
3	237.4	30.9	500	10	US-09-783-590-3621
4	219.8	28.6	379	10	US-09-960-352-154
5	214.8	27.9	400	10	US-09-960-352-156
6	190.8	24.8	335	10	US-09-783-590-3600
7	168.8	22.0	356	10	US-09-783-590-3575
8	133.2	17.3	288	10	US-09-783-590-3542
9	120.8	15.7	426	10	US-09-815-343-1490
10	106	13.8	1718	9	US-09-764-868-51
11	105.2	13.7	716	9	US-09-764-868-475
12	64.4	8.4	2686	9	US-09-764-868-48
13	64.4	8.4	3380	10	US-09-799-799-1
14	64	8.3	1091	9	US-09-764-868-316
15	43.4	5.6	3786	10	US-09-815-242-7865
16	42.4	5.5	1730	9	US-09-860-670-78
17	42.4	5.5	1730	9	US-09-764-868-52
18	42.4	5.5	1730	9	US-09-764-868-476
19	40.8	5.3	2172	10	US-09-815-242-4038

c	20	40.4	5.3	598	10	US-09-770-149-991	Sequence 991, App
	21	40.4	5.3	768	9	US-09-938-842A-812	Sequence 812, App
	22	39.6	5.1	4722	10	US-09-962-055-14	Sequence 14, Appl
	23	39.6	5.1	4722	12	US-10-023-529-14	Sequence 14, Appl
	24	39.6	5.1	4722	12	US-10-023-523-14	Sequence 14, Appl
c	25	39.4	5.1	420	10	US-09-960-352-9773	Sequence 9773, Ap
	26	38.4	5.0	1685	9	US-09-954-531-133	Sequence 133, App
	27	38.4	5.0	1685	9	US-09-954-531-354	Sequence 354, App
	28	38.4	5.0	1685	10	US-09-962-436-261	Sequence 261, App
	29	38.4	5.0	2296	10	US-09-822-849A-259	Sequence 259, App
c	30	37.8	4.9	420	10	US-09-960-352-9532	Sequence 9532, Ap
	31	37.8	4.9	422	9	US-09-854-133-337	Sequence 337, App
c	32	37.8	4.9	422	10	US-09-738-973-337	Sequence 337, App
	33	37.4	4.9	552	9	US-09-764-868-406	Sequence 406, App
	34	37.4	4.9	16747	10	US-09-764-877-3354	Sequence 3354, Ap
	35	36.6	4.8	273	10	US-09-864-761-17066	Sequence 17066, A
	36	36.6	4.8	403	10	US-09-960-352-11885	Sequence 11885, A
	37	36.6	4.8	1369	10	US-09-443-704-17	Sequence 17, Appl
	38	36.6	4.8	4610	9	US-09-884-001-1	Sequence 1, Appli
	39	36.2	4.7	866	10	US-09-764-870-262	Sequence 262, App
	40	36.2	4.7	3121	9	US-10-033-245-6	Sequence 6, Appli
	41	36.2	4.7	3121	9	US-10-033-223-6	Sequence 6, Appli
	42	36.2	4.7	3121	9	US-10-033-167-6	Sequence 6, Appli
	43	36.2	4.7	3121	12	US-10-033-246-6	Sequence 6, Appli
	44	36.2	4.7	3121	12	US-10-033-301-6	Sequence 6, Appli
	45	36.2	4.7	3121	12	US-10-033-326-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-925-297-220
; Sequence 220, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 220
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-220

Query Match	99.1%	Score 762;	DB 10;	Length 3094;
Best Local Similarity	100.0%;	Pred. No. 1.4e-216;		
Matches	762;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
Qy	8	GTGGTTTCAGAGCAGTGAGCAAGAGAGCCCATGCCGGAAGCACTGAAAGTCTCAT	67	
Db	62	GTGGTTTCAGAGCAGTGAGCAAGAGAGCCCATGCCGGAAGCACTGAAAGTCTCAT	121	
Qy	68	ATTCGCCGAATTTGACCTTGCACAAATTTCTACTACTATTTCTCAAGGAATTTGACCA	127	
Db	122	ATTCGCCGAATTTGACCTTGCACAAATTTCTACTACTATTTCTCAAGGAATTTGACCA	181	
Qy	128	ACGACTTGCCCTGTGGGAAGGCCGCTCAATGCCAAATCAGAGATTACCAAGAAATCGG	187	

```
|||||
Db 182 AGCACTTGCCCTGTGGAGGCGCTCAAAATGCCAAAATCAGAGATTACCAAGAAATCGG 241
QY 188 CGATGTCATGCTGAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCTGTGTGAAGCA 247
Db 242 CCATGTCTATGCTGAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCTGTGTGAAGCA 301
QY 248 CAGCAGGCGCTTGGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 307
Db 302 CAGCAGGCGCTTGGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 361
QY 308 CTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGTCAACACCTTCTCTCTCGG 367
Db 362 CTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGTCAACACCTTCTCTCTCGG 421
QY 368 GCACCTGCACCGGCTCATGCACTACAAGCAGTCTCTGAGCGGCTGTGCAACACCAACCC 427
Db 422 GCACCTGCACCGGCTCATGCACTACAAGCAGTCTCTGAGCGGCTGTGCAACACCAACCC 481
QY 428 GCCGAGCCACGCCGCTTCAGGAGCTGCCGAGCCGCTTTGGCAGAGATCACGGAGATGTT 487
Db 482 GCCGAGCCACGCCGCTTCAGGAGCTGCCGAGCCGCTTTGGCAGAGATCACGGAGATGTT 541
QY 488 GGCACAGCTCCACGTCAGATGATCAAGATGGAGAAATTTCCAGAGCTGCACGAACCTCAA 547
Db 542 GGCACAGCTCCACGTCAGATGATCAAGATGGAGAAATTTCCAGAGCTGCACGAACCTCAA 601
QY 548 GAAAGATTGATGTCATGACAACTCTTGTGTTCCGGGAAGGAGTTATCCGCTCTGGG 607
Db 602 GAAAGATTGATGTCATGACAACTCTTGTGTTCCGGGAAGGAGTTATCCGCTCTGGG 661
QY 608 CAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGA 667
Db 662 CAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGA 721
QY 668 CTTCTGCTATACAGAGCCGGGGCTGACGGCCCTCCAAATCAGTTTAAAGTCCAGGGCA 727
Db 722 CTTCTGCTATACAGAGCCGGGGCTGACGGCCCTCCAAATCAGTTTAAAGTCCAGGGCA 781
QY 728 GTCCTCCGCTCTATGTCATGACGATTGAGGAGCGAAGACGA 769
Db 782 GTCCTCCGCTCTATGTCATGACGATTGAGGAGCGAAGACGA 823
```

```
RESULT 2
US-09-783-590-3620
; Sequence 3620, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3620
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
```

```
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (340)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (352)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (356)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (399)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (402)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (406)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3620

Query Match 32.3%; Score 248.6; DB 10; Length 408;
Best Local Similarity 85.0%; Pred. No. 2.2e-64;
Matches 347; Conservative 0; Mismatches 46; Indels 15; Gaps 6;

QY 39 GCCATGCCGGAAGCACTCAAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTT 98
Db 1 GGCAGAGCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTT 60
QY 99 CATACTAATTTTCTCAAGGAAATTGAGCAAGCACTTGCCCTGTGGGAAGCGCGCTCAAAAT 158
Db 61 CATACTAATTTTNTCAAGGAAATTGAGCAAGCACTTGCCCTGTGGGAAGCGCGCTCAAAAT 120
QY 159 GCCAAATCAGAGATTACCAAGAAATCGCGATGTATGCTGGAAGAAATTCAGGGCATG 218
Db 121 GCCAAATCAGAGATTACCAAGAAATCGCGATGTATGCTGGAAGAAATTCAGGGCATG 180
QY 219 AAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGA-GGCTTGGAGGCCCTG---GAGAA 274
Db 181 AAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGAAGGAGGCCCTTGGAGGCCCTTGGAGGAAAT 240
QY 275 TGAATCAAGAGCTCCCGCGGCTGGAGA----ACTTCTGCAGAGACTTTTGAAGCTGCAGA 330
Db 241 GGAATCAAGAGCTCCCGCGGCTGGAGAACTTTCTGCCAGAGACTTTGAGCTGCAGA 300
QY 331 AGTGTGTATTACCTACCGCTCAACACCTTCTCTCTGCG-----GGCAGCTGCACCGGCTCAT 385
Db 301 AGTGTGTATTACCTAACCGNTTCAACAACTTTCTCTCTGNGGGCCACTGCANCNGNTCAT 360
QY 386 -GCACCTACAAGCAGTCTCTGGAGC-GGCTGTGCAAAACACCAACCGCGG 431
Db 361 GGCATTACAAGCAAGTTCTGGAGCGGTTGTGCAAAACANCAACNCG 408
```

```
RESULT 3
US-09-783-590-3621
; Sequence 3621, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
```

NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3621
LENGTH: 500

TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
LOCATION: (73)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (146)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (240)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (246)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (275)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (336)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (377)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (380)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (390)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (413)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (436)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (450)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (455)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (459)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (461)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (475)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (477)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (487)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (499)
OTHER INFORMATION: n equals a,t,g, or c

US-09-783-590-3621

Query Match 30.98; Score 237.4; DB 10; Length 500;
Best Local Similarity 83.34; Pred. No. 5.3e-61;
Matches 415; Conservative 0; Mismatches 60; Indels 23; Gaps 13;

Qy 39 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCGAATTTTGAACCTTTGCACAAATTT 98
Db 1 GGCAGAGCGGAAGCACTGAAAAGTCTCATATTCGCGAATTTTGAACCTTTGCACAAATTT 60
Qy 99 CATACTAATTTTCTCAAGGAAATTTGAGCAACGACTTGGCCCTGT-GGGAAGCGCGCTCAA 157
Db 61 CATACTAATTTTNTCAAGGAAATTTGAGCAACGACTTGGCCCTGTGGGGAAGCGCGCTCAA 120
Qy 158 TGCCCAATCAGAGATTACCAAGAATCGCGATGCTCATGCTGAA-GAACATTCAGGGCA 216
Db 121 TGCCCAATCAGAGATTACCAAGAATCGCGATGCTCATGCTGAAGGAACATTCAGGGCA 180
Qy 217 TGAAGCACTTGGCGG-CTCACCTGT-GGAAGCACAGCG-AGGCCCTTGGAGGCCCTGGAGA 273
Db 181 TGAAGCACTTGGCGTCTCACCTGTGGGAAGCACAGCGAAGGCCCTTGGAGGCCCTGGAGN 240
Qy 274 A-TGGAATCAAGAGCTCCCGCGG--CTGGAGAACTTCTGCAGAGACTTT-GAGCTGCA 328
Db 241 AATGGAATTCAGAGCTCCCGCGGCTGGAGGAGNCCTTCTGCAGAGACTTTGGAGCTGCA 300
Qy 329 GAAGGTGTG-TTACCTACCGCTCAACACCTT--CCTCCTCGCGCCACTGCACCGGCTCAT 385
Db 301 GAAGGTGTGTTTACCTACCGCTCAACACCTTCTCTNCTGCGGGCCACTGCACCGGTTTCA 360
Qy 386 GCACCTAACG--CAGGTCTCTGAGCGGCTGTGCACAAACACCCCGCGAGCCGCCGA- 442
Db 361 GCACCTAACGCGAGTCTCTGAGCGGNTTTCACAAACACCCCGCATCGNGSCCAAG 420
Qy 443 -----CTTCAGGAGCTCCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTGGCACACT 496
Db 421 GCCGATTTTTCAGGAGTCCCGAGCGG-TTNGCAGNGTTNANGGAGTTGTTGGGANANTT 479
Qy 497 CCAGGTACGATGATCAA 514
Db 480 CCAGGTNGTGTTCAA 497

RESULT 4

US-09-960-352-154

; Sequence 154; Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Bvatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 154
; LENGTH: 379
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E1-A9
US-09-960-352-154

Query Match 28.6%; Score 219.8; DB 10; Length 379;
Best Local Similarity 90.8%; Pred. No. 7.8e-56;
Matches 246; Conservative 0; Mismatches 22; Indels 3; Gaps 1;

Qy 1 TCACCTTCGTGGTTTCAGAGCACAGTGAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAA 60
Db 109 TCACCTTCGTGGTTTCAGAGCACAGTGAGCAAGAGGAGCTCCATGCCCGGAACCTTTGAAGA 168
Qy 61 GTCATATTCGCGAATTTTGAACCTTTGCACAAATTTTCATACATAATTTTCTCAAGGAAA 120
Db 169 GTCATATTCGCGAATTTTGAACCTTTGCACAAATTTTCACACAAATTTTCTCAAGGAAA 228
Qy 121 TTGAGCAACGACTTGGCCCTGTGGGAAGCGCGCTCAATGCCCAATCA---GAGATTACC 177
|||||

Db 229 TTGACCAACGACTTGGCCCTGTGGGAAGCCGCTCGAATGCCCATCATCAGAGGATTACC 288
 Qy 178 AAGAATCGGGATGTCATGCTGAAGAACATTTCAGGGCATGAAGCAGCTGGCGCTCACC 237
 Db 289 AGAATCGGAGATGTAATGCTGAAGAAATTCAGGGATGAAGCAACTGGCCGCCCACT 348
 Qy 238 TGTGGAAGCACAGCGAGCCCTTGGAGGCCCT 268
 Db 349 TGTGGAAGCACAGCGAGCGCTGGAGGGCT 379

RESULT 5
 US-09-960-352-156
 ; Sequence 156, Application US/09960352
 ; Patent No. US20020137139A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Warren, Wesley C.
 ; APPLICANT: Tao, Nengbing
 ; APPLICANT: Byatt, John C.
 ; APPLICANT: Mathialagan, Nagappan
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 ; FILE REFERENCE: 16511.006/37-21(10298)C
 ; CURRENT APPLICATION NUMBER: US/09/960,352
 ; CURRENT FILING DATE: 2001-09-24
 ; NUMBER OF SEQ ID NOS: 15112
 ; SEQ ID NO 156
 ; LENGTH: 400
 ; TYPE: DNA
 ; ORGANISM: Bos taurus
 ; OTHER INFORMATION: Clone ID: 01-LTB34-020-Q1-E2-A9
 US-09-960-352-156

Query Match 27.98; Score 214.8; DB 10; Length 400;
 Best Local Similarity 87.88; Pred. No. 2.5e-54;
 Matches 258; Conservative 0; Mismatches 32; Indels 4; Gaps 2;

Qy 1 TCACCTCGTGGTTTCAGAGCACAGTGGAGAGGAGCGCCATGCCGGAAGCACTGAAAA 60
 Db 107 TCACCTCGTGGTTTCAGAGCACAGTGGAGAGGAGCTCCATGCCGGAACCTTGAGAA 166
 Qy 61 GTCATATTCGCCGAATTTGAACCTTTGCACAAATTTCAATATTTCTCAAGGAAA 120
 Db 167 GTCATATTCGCCGAATTTGAACCTTTGCACAAATTTCAATATTTCTCAAGGACA 226
 Qy 121 TTGAGCAACGACTTGGCCCTGTGGGAAGCCGCTCAATGCCCAATCA---GAGATTACC 177
 Db 227 TTGAGCAACGACTTGGCCCTGTGGGAAGCCGCTCGAATGCCCATCATCAGAGGATTACC 286
 Qy 178 AAGAATCGGGATGTCATGCTGAAGAACATTTCAGGGCATGAAGCAGCTG-GCGGCTCAC 236
 Db 287 ATAGATCGGAGATGTACTGCTGAAGAACATTTCAGGGATGAAGCAACTGAGCTGCTCAC 346
 Qy 237 CTGTGGAAGCACAGCGAGCCCTTGGAGGCCCTGGAGATGGAATCAAGAGCTCC 290
 Db 347 TTGTGGAAGCACAGCTAGGCCCTTGGAGGCCGCTGGACATCGGCATCTCTGGGCCCC 400

RESULT 6
 US-09-783-590-3600
 ; Sequence 3600, Application US/09783590
 ; Patent No. US20020110850A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dillon, Patrick J.
 ; APPLICANT: Haseltine, William A.
 ; APPLICANT: Li, Haodong
 ; APPLICANT: Rosen, Craig A.
 ; APPLICANT: Ruben, Steven M.
 ; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
 ; FILE REFERENCE: PO-16.2C1
 ; CURRENT APPLICATION NUMBER: US/09/783,590
 ; CURRENT FILING DATE: 2000-02-15
 ; PRIOR APPLICATION NUMBER: 08/420,856

; PRIOR FILING DATE: 1995-04-12
 ; PRIOR APPLICATION NUMBER: 08/346,731
 ; PRIOR FILING DATE: 1994-11-21
 ; NUMBER OF SEQ ID NOS: 12485
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 3600
 ; LENGTH: 335
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (5)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (29)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (36)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (42)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (72)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (73)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (144)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (157)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (184)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (246)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (254)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (271)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (278)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (298)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (329)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (333)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-09-783-590-3600

Query Match 24.88; Score 190.8; DB 10; Length 335;
 Best Local Similarity 84.0%; Pred. No. 3.1e-47;
 Matches 278; Conservative 0; Mismatches 42; Indels 11; Gaps 6;

Qy 39 GCCATGCCGGAAGCACTGAAAGTCTCATATTCGCCGAATTTGAACTTTGACAAATTT 98
 Db 1 GGCANAGCGGAAGCACTGAAAGTCTCANATTCNAAATTTTAAACCTTTGCACAAATTT 60
 Qy 99 CATACTAATTTTCTCAAGGAATTTGACAGCACTTCCCTGTGGGAAGCCGCTCAAT 158
 Db 61 AATCTAATTTTNNCTCAAGGAATTTGACAGCACTTCCCTGTGGGAAGCCGCTCAAT 120
 Qy 159 GCCCAATCAG-AGATTACCAAGAATCGCGATGTCTGCTGAAGAACATTTCA-GGGCA 216

```
Db 121 GCCCAATCAGAAGATACCAAAATCGGCATGTATGCTGAAGAACATTCAGGGCA 180
Qy 217 TGAACACCTGGCGCTCACCTGTGGAAGCACAGGAGCGCTTGAGGCCCTGGAGAA-- 274
Db 181 TGANGACCTGGGGCTCACCTGTGGAAGCACAGGAGCGCTTGGAAGGCCCTGGAAGG 240
Qy 275 ---TGAATCAAGAGCTCCCGC-GGCTGGAGAACTTCTGCAGA-GACTTTGAGCTGC-- 327
Db 241 ATTGNATTCAAGTCTCCCGCGGCTGGNGAACTTNTGCAGAGGACTTTGAGCTGNCA 300
Qy 328 AGAAGGTGTGTACTACCGCTCAACACTT 358
Db 301 GAAAGGTGTGTACTACCGCTCAAAAANTT 331

RESULT 7
US-09-783-590-3575
; Sequence 3575, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3575
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (32)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (45)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (51)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (61)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (75)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (134)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (145)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (168)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (216)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (246)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (267)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (276)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (280)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (325)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (335)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (338)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3575

Query Match 22.0%; Score 168.8; DB 10; Length 356;
Best Local Similarity 77.9%; Pred. No. 1.le-40;
Matches 247; Conservative 0; Mismatches 57; Indels 13; Gaps 4;

Qy 39 GCCATGCGGGAAGCACTGAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTT 98
Db 1 GGCAGAGCGGAAGCACTGAAAGTCTCATATNCCGAATTTNAANCCTTTNCACAAATTT 60
Qy 99 CATACTAATTTTCTCAAGAAATTTGACAACGACTTGCCCTGTGGGAAGCGCGCTCAAA 158
Db 61 NATACTAATTTTNTNAAGAAATTAAGCAACGACTTGCCCTGTGGAAAGCGCGCTCAAA 120
Qy 159 GCCAATATCAGAGATTACCAAGATCGCGCATGT-CATGCTGAAGACATTCAGGGCAT 217
Db 121 GCCCAATCAGAGNTTACCAAGANTCGGGATGTGCATGTGAAGANCAATTCAGGGCAT 180
Qy 218 GAAGCACCTGGCGCTCACCTGT-GGAAGCACAGGAGCGCTTGAGGCGCTTGGAAGATG 276
Db 181 GAGCACCTGGGGTTTACCTGTGGAAGCACAGCNAGGCGCTTGAAGCGCTTGAAGG 240
Qy 277 GAATCAA-----GAGTCCCGGGGCTGGAGAACTTCTG-----CAGAGACTTTGAGCT 325
Db 241 AATGNATTCAAGAGGTTCGCGGGGNTGGAGGAATTTTNTGCCAGAGAACTTTGAAGCT 300
Qy 326 GCAGAGGTGTGTACC 342
Db 301 GCAGAGGTGTGTAAAC 317

RESULT 8
US-09-783-590-3542
; Sequence 3542, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
```

;; PRIOR FILING DATE: 1994-11-21
;; NUMBER OF SEQ ID NOS: 12485
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 3542
;; LENGTH: 288
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (3)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (42)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (72)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (75)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (93)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (129)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (142)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (155)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (156)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (166)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (182)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (235)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (263)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (264)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (267)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (270)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (272)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (282)
;; OTHER INFORMATION: n equals a,t,g, or c
;; US-09-783-590-3542

Query Match 17.3%; Score 133.2; DB 10; Length 288;
Best Local Similarity 85.88; Pred. No. 3.8e-30;
Matches 188; Conservative 0; Mismatches 27; Indels 4; Caps 4;
QY 46 CGGAAGCACTGAAAAGTCTCAATTTCCCAATTTTGAACCTTTTCACAAATTTTCACTACTA 105
|||||
; Sequence 51, Application US/09764868
; Patent No. US20020168711A1

Db 8 CGGAAGCACTGAAAAGTCTCAATTTCCCAATTTTAAACCTTTTCACAAATTTTAAACTA 67
QY 106 ATTTTCTCAAGGAAATTTGAGCAACGACTTGCCTTGTGGGAAGCGCTCAATGCCCCAAA 165
|||||
Db 68 ATTTNTTNAAGGAAATTTAAGCAACGCTTGCCTTGTGGAAAGCGCTCAATGCCCCAAA 127
QY 166 TCAGAGATTACCAAAGAAATGGCGATGT-CATGCTGAAG-AACATTTCAGGCGCATGAAGCA 223
|||||
Db 128 TNA-AGATTACCAANAATCGCGATGTNNATGCTGAAGNANCATTTCAGGCGCATGNGCA 186
QY 224 CTTGGCGGC-TCACCTCTGGAAGCACAGCGAGCGCTTGG 261
|||||
Db 187 CTTGGCGGCTTCACTCTGGAAGCACAGCGAGCGCCTTG 225
RESULT 9
US-09-815-343-1490/c
; Sequence 1490, Application US/09815343
; Patent No. US2001005596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121-504
; CURRENT APPLICATION NUMBER: US/09/815,343
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1490
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-815-343-1490

Query Match 15.7%; Score 120.8; DB 10; Length 426;
Best Local Similarity 60.2%; Pred. No. 2.3e-26;
Matches 219; Conservative 0; Mismatches 142; Indels 3; Gaps 1;
QY 1 TCACCTCTGCTGTTTCAGAGCACAGTGAAGAGGAGCGCATGCCGGAAGCACTGAAAA 60
|||||
Db 365 TTACCGTGTGTTCCGACGCGCACTGTTGAAGGAGGAGCGCATGCCCTGCGACTCTGATGA 306
QY 61 GTCTCATATTCCCGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAAA 120
|||||
Db 305 CGTGTCTTCTTCCACATCATCCCATATGATTTCCACAGAGGCTTCTCGCGGAGG 246
QY 121 TTGAGCAACGACTTGCCTGTGGGAAGGCGCTCAATGCCCAATCAGAG---ATTACC 177
|||||
Db 245 TGGAGCAGAGGCTGGCACTCTGGGAAGGCGCTCCAAAGCCACACAAAAGGCGAGTCATC 186
QY 178 AAAGAATCGGCGATGTCATGCTGAAGAACATTTTCAGGCGCATGAAGCACTGGCGGCTCACC 237
|||||
Db 185 AACGAATCGGGGACATCCTGCTCAGGAACATCGCCAGTAAAGGAGTTTACCAGCTACT 126
QY 238 TGTGGAAGCACAGCGAGCGCTTGGAGCGCTTGGAGATGGAATCAAGAGCTCCCGGGCGG 297
|||||
Db 125 TCCAAAGACATGAGAGATCCTTAACAGAACTGGAAGAGGCTACCAACGCTGTGAAGAGT 66
QY 298 TGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACTACCTACCGCTCAACACCT 357
|||||
Db 65 TGGAGGCACTGTACAGAGAGTTTGAGCTGCGAAGGCTGCGAAGAGTCTGCTACTTGCCTCAACAGCT 6
QY 358 TCCT 361
|||||
Db 5 TCCT 2

RESULT 10
US-09-764-868-51
; Sequence 51, Application US/09764868
; Patent No. US20020168711A1


```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-51

Query Match      13.8%; Score 106; DB 9; Length 1718;
Best Local Similarity 64.9%; Pred. No. 1.3e-21;
Matches 157; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 528 CAGAAGCTGCACGAACTCAAGAAAGATTGATTGGCATTGACAATCTTGTGTTCCGGGA 587
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13 CAGAAGCTAAAGAGCTGCAGCGGACCTGGTGGCATAGAGAACCTCATTCTCTGGC 72

QY 588 AGGGAGTTTCCTGCTGGGACGCTCAGCAAGCTCTCGGGGAAGGGCTCAGCAGGCG 647
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73 AGGGAGTTTCCTGCTGGGAGCTGCTTCAAAAGCTCACCAGAAAGGGCTCGCAGCAGAG 132

QY 648 ATGTCTTCTCTTCAACGACGCTCTGCTATACACGAGCGGGGCTGACGCCCTCCAAAT 707
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 ATGTTTTTCTCTTCAGATATGTTGCTGTACACAAGCAAGAGGATGTCAGGGACCCAGC 192

QY 708 CAGTTAAAGTCCACGGGAGCTCCCGCTCTATGCATGACGATTGAGGAGCGGAAGAC 767
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193 CACTTCCGGATCCGGGGCTCTTCCCTTCCCAAGGATGCTGGTGAAGAAAGTGATAAC 252

QY 768 GA 769
      ||
Db 253 GA 254

RESULT 11
US-09-764-868-475
; Sequence 475, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 475
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (715)
; OTHER INFORMATION: n equals a,t,g, or c

Query Match      13.7%; Score 105.2; DB 9; Length 716;
Best Local Similarity 64.0%; Pred. No. 1.3e-21;
Matches 155; Conservative 2; Mismatches 85; Indels 0; Gaps 0;

QY 528 CAGAAGCTGCACGAACTCAAGAAAGATTGATTGGCATTGACAATCTTGTGTTCCGGGA 587
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 CAGAAGCTWACGGAGCTGCKWKGCGGACCTGGTGGCATAGAGAACCTCATTGCTCTGGC 62

QY 588 AGGGAGTTTCCTGCTGGGACGCTCAGCAAGCTCTCGGGGAAGGGCTCAGCAGGCG 647
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 63 AGGGAGTTTCATCCGTGAGGGCTGCTTCAAAAGCTCACCAGAAAGGGCTCGCAGCAGAGG 122
QY 648 ATGTCTTCTCTTCAACGACGCTCTGCTATACACGAGCGGGGCTGACGGCTCCCAAT 707
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 ATGTTTTTCTCTTTCAGATATGTTGCTGTACACAAGCAAGAGGATGTCAGGGACCCAGC 182

QY 708 CAGTTAAAGTCCACGGGAGCTCCCGCTCTATGCATGACGATTGAGGAGCGGAAGAC 767
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 CACTTCCGGATCCGGGGCTCTTCCCTTCCCAAGGATGCTGGTGAAGAAAGTGATAAC 242

QY 768 GA 769
      ||
Db 243 GA 244

RESULT 12
US-09-764-868-48
; Sequence 48, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 2686
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (2505)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2569)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2644)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2645)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-48

Query Match      8.4%; Score 64.4; DB 9; Length 2686;
Best Local Similarity 47.8%; Pred. No. 3.9e-09;
Matches 287; Conservative 0; Mismatches 301; Indels 12; Gaps 3;

QY 175 ACCAAGAATCGGCGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGCGGCTC 234
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1505 ACCCAGCGCTCGGAGACATCTCTGAGAAAGCTGGCCCATCTCTGAAGATGTACGGCGAGT 1564

QY 235 ACCTGTGGAAGCACAGCGAGGCTTGGAGCGCTTGAGAGATGGAATCAAGAGCTCCCGGC 294
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1565 ATGCTCAAGAACTTTGACCCAGCGCTGAGGCTGCTGAGCACGTGGACCCAGCGCTCCCCAC 1624

QY 295 GGCTGGAGAACTTCTGACAGAGACTTTGAGCTGCAGAAAGGTGT---GTTACTTACCGCTCA 351
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1625 TGTTTAAAGACGCTCTCCACAGCATCCAGAGCAGGAGGTATGCGGGAACCTTGACGCTGC 1684

QY 352 ACACCTTCTCTCTCGGCGCACTGCACCGGCTCATGCATACAGCAGGTCCTTGGAGCGGC 411
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1685 AGCACACATGCTGGAGCGCGTGCAGAGGGTCCCGGCTACGAGCTGTGCTCAAGACT 1744

QY 412 TGTGCAAAACACACCGCGAGCGCTTCAGGGACTGCGGAGCGCTTTGGCAG 471
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1745 ATCTGAAGAGGCTCCCGCAGGAGCGCCACAGCCGGAAGGATCGCGAGAGGCTCTTTGGAGC 1804

QY 472 AGATCAGCGAGATGTTGGCAGAGCTCCAGGTTACGATGATCAAGATGGAAGATTTCCAGA 531
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1805 TCATCTCCACAGCCGCCAACCACTCCAATGTCGCAATTCGGAAAGTGAGAAAATGCACA 1864
QY 532 AGCTGCAGCAACTCAAGAAAGATTTGATTGGCATTTGACAAATCTTGCTGGTTCGGGAAGGG 591
Db 1865 ACCTTTGGAGGTGTACGAGAGCTGGGTGGGAAGAGACATTTGTCAACCGGCCAATG 1924
QY 592 AGTTTCATCGTCTGGCAGCCCTCAGCAAGCTCTCGGGGAAG-----GGGCTCCAGCAGC 645
Db 1925 AACTGATCAAGGAGGGCCAAATCCAGAAACTGTACGCCAAGAACGCCACCCCCAGGACC 1984
QY 646 GCATGTTCTTCTTTCACAGAGCTCTGCTATACAGAGCCGGGGGCTGACGGCC---T 702
Db 1985 GCACCTCTTCTGTTCAACAGCATGATCTTTACTGTGTGCCCAAGTGGCGCTCATGG 2044
QY 703 CCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAGAGCG 762
Db 2045 GCAGAAGTTCAGCTCGGGAGAAGATGGACATCTCAGGCCCTCAGGTGTCAGGATATCG 2104

RESULT 13

US-09-799-799-1
; Sequence 1, Application US/09799799
; Patent No. US20020132291A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001157
; CURRENT APPLICATION NUMBER: US/09/799,799
; CURRENT FILING DATE: 2001-03-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3380
; TYPE: DNA
; ORGANISM: Human
US-09-799-799-1

Query Match 8.4%; Score 64.4; DB 10; Length 3380;
Best Local Similarity 47.8%; Pred. No. 4.e-09;
Matches 287; Conservative 0; Mismatches 301; Indels 12; Gaps 3;

QY 175 ACCAAGAATCGCGCATGTCATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGCTC 234
Db 1225 ACCCAGGCTCGGGACATCTCTGAGAGCTGGCCCCATCTCTGAGATGTACGCGGAGT 1284
QY 235 ACCTGTGGAAGCACAGCGAGGCTTTGGAGGCCCTTGGAATGGAATCAAGAGCTCCCGGC 294
Db 1285 ATGTCAAGAACTTTGACCGAGCGGTAGGCTGGTGAGCACGTGGACCCAGCGCTCCCCAC 1344
QY 295 GGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTGT--GTTACCTACCGTCA 351
Db 1345 TGTTTAAAGACCTGTCACAGCATCCAGAGCAGGAGGTATGCGGGAACCTTCAGCGTGC 1404
QY 352 ACACCTTCTCTCGGGGCACCTGCACCGGCTCATGCATCTACAAGCAGGTCTCTGGAGCGGC 411
Db 1405 AGCACCATGCTGAGCGCGTGAGAGGGTCCCGCGTACGAGCTGTGCTCAAGGACT 1464
QY 412 TGTGAACACACACCGCGAGCCACGCGGACTTCAGGAGCTTCGCGAGCCGCTTTGGCAG 471
Db 1465 ATCTGAAGAGGCTCCCGAGGAGCGCCAGACCGGAAGGTGCGGAGAGGTCTCTGGAGC 1524
QY 472 AGATCAGGAGATGTGGCAGAGCTCCACGCTACCATGATCAAGATGAGAGATTTCCAGA 531
Db 1525 TCATCTCCACAGCCGCCAACCACTCCAATGTGCGCATTCGGAAAGTGGAGAAAATGCACA 1584
QY 532 AGCTCAGCAACTCAAGAAAGATTTGATTGGCATTTGACAATCTTGCTGGTTCGGGAAGGG 591
Db 1585 ACCTTTGGAGGTGTACGAGCAGCTGGGTGGGAAGAGACATTTGTCAACCGGCCAATG 1644
QY 592 AGTTTCATCGTCTGGCAGCCTCAGCAAGCTCTCGGGGAAG-----GGGCTCCAGCAGC 645
Db 1645 ATCTGAAGAGGCTCCCGAGGAGCGCCAGACCGGAAGGTGCGGAGAGGTCTCTGGAGC 1524

Db 1645 AACTGATCAAGGAGGGCCAAATCCAGAAACTGTACGCCAAGAACGGACCCCCAGGACC 1704
QY 646 GCATGTTCTTCTGTTCAACGAGCTCTGCTATACAGAGCGGGGGCTGACGGCC---T 702
Db 1705 GCCACCTCTTCTGTTCAACGAGCATGATCTTTACTGTGTCGCCAAGCTGGCGCTCATGG 1764
QY 703 CCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAGAGCG 762
Db 1765 GCAGAAGTTCAGCTCGGGAGAAGATGGACATCTCAGGCCCTCAGGTGTCAGGATATCG 1824

RESULT 14

US-09-764-868-316
; Sequence 316, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 316
; LENGTH: 1091
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1059)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-316

Query Match 8.3%; Score 64; DB 9; Length 1091;

Best Local Similarity 47.7%; Pred. No. 3.le-09;

Matches 286; Conservative 1; Mismatches 301; Indels 12; Gaps 3;

QY 175 ACCAAGAATCGCGCATGTCATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGCTC 234
Db 188 ACCCAGGCTCGGGACATCTCTGAGAGCTGGCCCCATCTCTGAGATGTACGCGGART 247
QY 235 ACCTGTGGAAGCACAGCGAGGCTTTGGAGGCCCTTGGAATGGAATCAAGAGCTCCCGGC 294
Db 248 ATGTCAAGAACTTTGACCGAGCGGTAGGCTGGTGAGCACGTGGACCCAGCGCTCCCCAC 307
QY 295 GGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGAGGTGT---GTTACCTACCGTCA 351
Db 308 TGTTTAAAGAGCTCTCCACAGCATCCAGAGCAGGAGGTATGCGGGAACCTTCAGCGTGC 367
QY 352 ACACCTTCTCTCGGGGCACCTGCACCGGCTCATGCATCTACAAGCAGGTCTCTGGAGCGGC 411
Db 368 AGCACCATGCTGGAGCGCGTGACAGAGGTCCCGCGTACGAGCTGTGCTCAAGGACT 427
QY 412 TGTGAACACACACCGCGGAGCCACGCGGACTTCAGGAGCTGCGGAGCGGCTTTGGCAG 471
Db 428 ATCTGAAGAGGCTCCCGAGGAGCGCCCGAGACGGAAGATGCGGAGAGGTCTTTGGAGC 487
QY 472 AGATCAGGAGATGTGGCAGAGCTCCACGCTACGATGATCAAGATGAGAGATTTCCAGA 531
Db 488 TCATCTCCACAGCCGCCAACCACTCCAATGTGCTGCGCATTCGGAAAGTGGAGAAAATGCACA 547
QY 532 AGCTCAGCAACTCAAGAAAGATTTGATTGGCATTTGACAATCTTGCTGGTTCGGGAAGGG 591
Db 548 AGCTTTGGAGGTGTACGAGCAGCTGGGTGGGAAGAGACATTTGTCAACCGGCCAATG 607
QY 592 AGTTTCATCGTCTGGCAGCCTCAGCAAGCTCTCGGGGAAG-----GGGCTCCAGCAGC 645
Db 608 AACTGATCAAGGAGGCCAAATCCAGAAACTGTACGCCAAGAACGCCACCCCCAGGACC 667
QY 646 GCATGTTCTTCTGTTCAACGAGCTCTGCTATACAGAGCGGGGGCTGACGGCC---T 702
Db 668 GCCACCTCTTCTGTTCAACAGCATGATCTTTACTGTGTGCCCAAGCTGCGGCTCATGG 727

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 04:24:01 ; Search time 2433 Seconds
(without alignments)
5118.919 Million cell updates/sec

Title: US-09-555-342B-1_COPY_1733_2501

Perfect score: 769

Sequence: 1 tcactctgtggttcagagc.....attgaggagagcgaagacga 769

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_Other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	744.8	96.9	1043	BQ072025	BQ072025 AGENCOURT
2	699.6	91.0	926	BQ706499	BQ706499 AGENCOURT
3	682	88.7	988	BQ674703	BQ674703 AGENCOURT
4	628.4	81.7	909	BQ646351	BQ646351 AGENCOURT
5	592	77.0	621	BG770181	BG770181 602744940
6	591	76.9	861	BG764061	BG764061 602737071

7	577.6	75.1	890	12	BE907778	BE907778
8	573.2	74.5	901	12	BG475554	BG475554
9	560.6	72.9	788	12	BG750463	BG750463
c 10	535.2	68.3	618	14	BQ331885	BQ331885
11	505.6	65.7	923	14	BQ950768	BQ950768
12	494.4	64.3	496	14	BM705217	BM705217
13	475.4	61.8	489	9	AL121548	AL121548
14	463	60.2	861	12	BG420356	BG420356
15	457	59.4	1026	13	BM551766	BM551766
16	444	57.7	454	14	BM728340	BM728340
17	414.8	53.9	896	14	BQ672632	BQ672632
18	412.4	53.6	757	10	BE282962	BE282962
19	398	51.8	719	12	BG248068	BG248068
20	366	47.6	449	10	BB840008	BB840008
21	340.4	44.3	741	12	BG122769	BG122769
22	337.4	43.9	351	14	D81819	D81819
c 23	336.8	43.8	358	12	BE812222	BE812222
24	326.8	42.5	466	14	RI4694	RI4694
25	323	42.0	516	13	BI898950	BI898950
26	322.6	42.0	475	14	BQ301106	BQ301106
c 27	314.6	40.9	876	14	BQ221203	BQ221203
28	304	39.5	414	10	BE271866	BE271866
29	304	39.5	516	13	BI535640	BI535640
30	303	39.4	1122	12	BE746268	BE746268
31	300.8	39.1	994	13	BI081073	BI081073
32	297.8	38.7	389	14	H11133	H11133
33	288.6	37.5	703	13	BJ059899	BJ059899
34	276	35.9	340	12	BF226117	BF226117
35	273.2	35.5	435	9	AI298217	AI298217
36	273.2	35.5	472	10	BE673853	BE673853
37	273.2	35.5	495	9	AI290936	AI290936
38	270	35.1	490	12	BF447678	BF447678
39	270	35.1	890	14	BQ681261	BQ681261
40	270	35.1	952	14	BQ676994	BQ676994
41	270	35.1	959	14	BQ679905	BQ679905
42	270	35.1	970	14	BQ676703	BQ676703
43	269.4	35.0	446	9	AI701115	AI701115
44	263.6	34.3	994	14	BQ683731	BQ683731
45	260	33.8	304	13	BI051634	BI051634

ALIGNMENTS

RESULT 1
BQ072025
LOCUS
DEFINITION BQ072025 AGENCOURT_6859787 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5928511
5', mRNA sequence.
ACCESSION BQ072025
VERSION BQ072025.1 GI:19901071
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2105 row: c column: 08
High quality sequence stop: 626.
Location/Qualifiers
1. .1043

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5928511"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 266 a 290 c 294 g 193 t
ORIGIN
Query Match 96.9%; Score 744.8; DB 14; Length 1043;
Best Local Similarity 99.7%; Pred. No. 4.7e-186;
Matches 746; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 9 TGGTTTCAGAGCACAGTGGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAAGTCTCAT 68
DB 1 TGGTTTCAGAGCACAGTGGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAAGTCTCAT 60
QY 69 TTCCCGAATTTTGAACCTTTGACAAAATTTTCATCTAATTTTCTCAAGGAATTCAGCAA 128
DB 61 TTCCCGAATTTTGAACCTTTGACAAAATTTTCATCTAATTTTCTCAAGGAATTCAGCAA 120
QY 129 CGACTTGGCCCTGTGGGAAGGCGCTCAATGCCCAATCAGAGATTACCAAGAATCGGC 188
DB 121 CGACTTGGCCCTGTGGGAAGGCGCTCAATGCCCAATCAGAGATTACCAAGAATCGGC 180
QY 189 GATGTCATGCTCAAGAACATTCAGGGCATGAGCACTTGGGGCTCACCTGTGGGAAGCAC 248
DB 181 GATGTCATGCTCAAGAACATTCAGGGCATGAGCACTTGGGGCTCACCTGTGGGAAGCAC 240
QY 249 ACGGAGGCTTGGAGGCGCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 308
DB 241 ACGGAGGCTTGGAGGCGCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 300
QY 309 TCGAGAGACTTTGAGCTGCAAGAGTGTGTACCTTACCGCTCAACACCTTCTCTCTGGG 368
DB 301 TCGAGAGACTTTGAGCTGCAAGAGTGTGTACCTTACCGCTCAACACCTTCTCTCTGGG 360
QY 369 CCAGTGCACCGGCTCATGCATCAGCAAGCAGTCTCGGAGCGGCTGCAAAACACCAACCG 428
DB 361 CCAGTGCACCGGCTCATGCATCAGCAAGCAGTCTCGGAGCGGCTGCAAAACACCAACCG 420
QY 429 CCGAGCCAGCGGCTTTCAGGAGCTGCGGAGCGGCTTTGGCAGAGATCAGCAGATGGTG 488
DB 421 CCGAGCCAGCGGCTTTCAGGAGCTGCGGAGCGGCTTTGGCAGAGATCAGCAGATGGTG 480
QY 489 GCACAGCTCCAGGTCAGATGATCAAGATGGAGAAATTTCCAGAGCTGCACGAACCTCAAG 548
DB 481 GCACAGCTCCAGGTCAGATGATCAAGATGGAGAAATTTCCAGAGCTGCACGAACCTCAAG 540
QY 549 AAAGATTTGATTTGACATTTGACAAATCTTGTGTTCCGGGAAGGAGTTTATCGCTGGGC 608
DB 541 AAAGATTTGATTTGACATTTGACAAATCTTGTGTTCCGGGAAGGAGTTTATCGCTGGGC 600
QY 609 AGCCTTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGAC 668
DB 601 AGCCTTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGAC 660
QY 669 GTCTGCTATACAGAGCGGGGCTGACGGCTTCAATCAGTTTAAAGTCCACGGGCGAG 728
DB 661 GTCTGCTATACAGAGCGGGGCTGACGGCTTCAATCAGTTTAAAGTCCACGGGCGAG 720
QY 729 CTCGGCTCTATGGCATGACGATTGAGG 756
DB 721 CTCGGCTCTATGGCATGACGATTGAGG 748
```

RESULT 2
LOCUS B0706499 926 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8474935 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301634
5', mRNA sequence.
ACCESSION B0706499
VERSION B0706499.1 GI:21845398
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 926)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rcgabs-r@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Cloned into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 242 a 247 c 254 g 183 t
ORIGIN

Query Match 91.0%; Score 699.6; DB 14; Length 926;
Best Local Similarity 98.6%; Pred. No. 4e-174;
Matches 716; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1 TCACCTTCGTGCTTTCAGAGCACAGTGGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAA 60
DB 114 TCACCTTCGTGCTTTCAGAGCACAGTGGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAA 173
QY 61 GTCATCATATCCCGAATTTTGAACCTTTGCACAAATTTTCTACTAATTTTCTCAAGGAAA 120
DB 174 GTCATCATATCCCGAATTTTGAACCTTTGCACAAATTTTCTACTAATTTTCTCAAGGAAA 233
QY 121 TTGAGCAACAGCACTTGCCTGTGGGAAGGCGCTCAATGCCCCAAATCAGAGATTACCAAAA 180
DB 234 TTGAGCAACAGCACTTGCCTGTGGGAAGGCGCTCAATGCCCCAAATCAGAGATTACCAAAA 293
QY 181 GAATCGCGGATGTCATGCTGAAGAACATTCAGGCGCATGAAGCACTGGCGCTCACCTGT 240
DB 294 GAATCGCGGATGTCATGCTGAAGAACATTCAGGCGCATGAAGCACTGGCGCTCACCTGT 353
QY 241 GGAAGCAGCAGCGGCGCTTGGAGGCGCTTGGAGATGGAATCAAGAGCTCCCGCGGCTGG 300
DB 354 GGAAGCAGCAGCGGCGCTTGGAGGCGCTTGGAGATGGAATCAAGAGCTCCCGCGGCTGG 413
QY 301 AGAATTTCTGAGAGACTTTGAGCTGCAAGAGGTTGTGTACCTACCGCTCAACACCTTCC 360
DB 301 AGAATTTCTGAGAGACTTTGAGCTGCAAGAGGTTGTGTACCTACCGCTCAACACCTTCC 360

Db 414 AGAAGTCTGCGAGACATTTGAGCTGCGAGAGGTGTGTTACCTACCGCTCAACACCTTCC 473
QY 361 TCCTGCGGCACCTGACCGGCTCATGCACTACAAGCAGGTCTCTGGAGCGGTGTGCAAAAC 420
Db 474 TCCTGCGGCACCTGACCGGCTCATGCACTACAAGCAGGTCTCTGGAGCGGTGTGCAAAAC 533
QY 421 ACCACCGCGAGCGACCGCGACTTCAGGAGCTCCGAGCGCTTGGCAGAGATCACGG 480
Db 534 ACCACCGCGAGCGACCGCGACTTCAGGAGCTCCGAGCGCTTGGCAGAGATCACGG 593
QY 481 AGATGTCGACAGCTCCACGCTACGATCATCAAGATGAGATTTCCAGAGCTGCACG 540
Db 594 AGATGTCGACAGCTCCACGCTACGATCATCAAGATGAGATTTCCAGAGCTGCACG 653
QY 541 AACTCAAGAAAGATTTGATTTGGCATTTGACATCTTTGTGTTCCGGGAAGGGAGTTTCATCC 600
Db 654 AACTCAAGAAAGATTTGATTTGGCATTTGACATCTTTGTGTTCCGGGAAGGGAGTTTCATCC 713
QY 601 GTCTGGCAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTTCTTCTGT 660
Db 714 GTCTGGCAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTTCTTCTGT 773
QY 661 TCAACGAGCTCTGCTATACAGAG - CCGGGGCTGACGGCTCCAAATCAGTTTAAAGTC 719
Db 774 TTCAGCAGCTCTGCTATACAGAGCGCGGGGTGGACGCCCTCCAATCAGTTTAAAGTC 833
QY 720 CACGGG 725
Db 834 CACGGG 839

RESULT 3
BO674703
LOCUS
DEFINITION BO674703 988 bp mRNA linear EST 15-JUL-2002
AGENCOURT_8188989 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255217
5', mRNA sequence.

ACCESSION BO674703
VERSION BO674703.1 GI:21785537
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Cloned by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2405 row: p column: 02
High quality sequence stop: 685.
Location/Qualifiers
1..988
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6255217"
/clone_lib="NIH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: salivary gland; Vector: pOTB7; Site:1; XhoI;
Site:2; EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGCAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life

BASE COUNT 216 a 306 c 291 g 161 t 14 others
Technology). Note: this is a NIH_MGC Library."

Query Match 88.7%; Score 682; DB 14; Length 988;
Best Local Similarity 99.6%; Pred. No. 1.8e-169;
Matches 693; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 75 AATTTTGAACCTTTTCACAAATTTTCATATAATTTTCAAGGAAATTTGAGCAACGACTT 134
Db 1 AATTTTGAACCTTTTCACAAATTTTCATATAATTTTCAAGGAAATTTGAGCAACGACTT 60
QY 135 GCCCTGTGGGAAGCGCTCAAAATCCCAAAATCAGAGATTTACCAAGAAATCGCGATGTC 194
Db 61 GCCCTGTGGGAAGCGCTCAAAATCCCAAAATCAGAGATTTACCAAGAAATCGCGATGTC 120
QY 195 ATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGCGGCTCACCTGTGGNAGCACAGCGAG 254
Db 121 ATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGCGGCTCACCTGTGGNAGCACAGCGAG 180
QY 255 GCCTTTGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGGGGCTGGAGAACTTCTGCGAGA 314
Db 181 GCCTTTGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGGGGCTGGAGAACTTCTGCGAGA 240
QY 315 GACTTTGAGCTGSCAGAAGGTGTGTTACCTACCGCTCAACACCTTCTCTGCGGCGCACTG 374
Db 241 GACTTTGAGCTGSCAGAAGGTGTGTTACCTACCGCTCAACACCTTCTCTGCGGCGCACTG 300
QY 375 CACCGGCTCATGCACTACAAGCAGGTCTCTGGAGCGGTGTGCAAAACACACCGCGGAGC 434
Db 301 CACCGGCTCATGCACTACAAGCAGGTCTCTGGAGCGGTGTGCAAAACACACCGCGGAGC 360
QY 435 CACCGGCTCATGCACTACAAGCAGGTCTCTGGAGCGGTGTGCAAAACACACCGCGGAGC 494
Db 361 CACCGGCTCATGCACTACAAGCAGGTCTCTGGAGCGGTGTGCAAAACACACCGCGGAGC 420
QY 495 CTCCACGCTACGATGATCAAGATGGAATTTCCAGAGCTGACAGAACTCAAGAAAGAT 554
Db 421 CTCCACGCTACGATGATCAAGATGGAATTTCCAGAGCTGACAGAACTCAAGAAAGAT 480
QY 555 TTGATTTGCAATTCAGAACTTCTGTTCCGGAAGGGAGTTTCATCCGTTGGGCGAGCTC 614
Db 481 TTGATTTGCAATTCAGAACTTCTGTTCCGGAAGGGAGTTTCATCCGTTGGGCGAGCTC 540
QY 615 AGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGCTCTG 674
Db 541 AGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGCTCTG 600
QY 675 CTATACACGAG - CCGGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCAGGGCAGCTCCC 733
Db 601 CTATACACGAGNCCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCAGGGCAGCTCCC 660
QY 734 GCTCTATGGCATGACGATTTGAGGAGCGGAGACGCA 769
Db 661 GCTCTATGGCATGACGATTTGAGGAGCGGAGACGCA 696

RESULT 4
BO646351
LOCUS
DEFINITION BO646351 909 bp mRNA linear EST 15-JUL-2002
AGENCOURT_8286354 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6299462
5', mRNA sequence.

ACCESSION BO646351
VERSION BO646351.1 GI:21770523
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 909)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)


```
QY 418 AACACCACCGCGGAGCCAGCGGACTTCAGGAGACTGCGAGCGGCTTTGGCAGAGATCA 477
|||||
Db 242 AACACCACCGCGGAGCCAGCGGACTTCAGGAGACTGCGGAGCGGCTTTGGCAGAGATCA 301
QY 478 CGGAGATGTTGGCAGAGCTCCACGGTACGATGATCAAGATGAGAGATTTCCAGAAGCTGC 537
|||||
Db 302 CGGAGATGTTGGCAGAGCTCCACGGTACGATGATCAAGATGAGAGATTTCCAGAAGCTGC 361
QY 538 ACGAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGTTCCGGGAAGGAGTTCA 597
|||||
Db 362 ACGAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGTTCCGGGAAGGAGTTCA 421
QY 598 TCCGTCGCGGAGCCCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 657
|||||
Db 422 TCCGTCGCGGAGCCCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 481
QY 658 TGTTCACAGAGCTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAG 717
|||||
Db 482 TGTTCACAGAGCTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAG 541
QY 718 TCCACGGGAGCTCCCGCTTATGGCATGACGATTGAGGAGAGCGAAGACGA 769
|||||
Db 542 TCCACGGGAGCTCCCGCTTATGGCATGACGATTGAGGAGAGCGAAGACGA 593

RESULT 6
Bg764061 861 bp mRNA linear EST 15-MAY-2001
LOCUS 602737071F1 NTH_MGC_49 Homo sapiens cDNA clone IMAGE:4862417 5',
DEFINITION mRNA sequence.
ACCESSION Bg764061
VERSION Bg764061.1 GI:14074714
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 861)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1722 row: n column: 18
High quality sequence stop: 817.
Location/Qualifiers
1. .861
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4862417"
/clone_lib="NTH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NTH_MGC
Library."
198 a 245 c 270 g 147 t 1 others

BASE COUNT
ORIGIN
```

```
Query Match 76.9%; Score 591; DB 12; Length 861;
Best Local Similarity 99.8%; Pred. No. 2e-145;
Matches 591; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 178 AAAGAATCGCGGATGCTATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGGCTCACC 237
|||||
Db 2 AAAGAATCGCGGATGCTATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGGCTCACC 61
QY 238 TGTGGGAAGCACAGCAGCGGCTTGGAGGGCCTTGAGAGATGAATCAAGAGTCCCGCGGCG 297
|||||
Db 62 TGTGGGAAGCACAGCAGCGGCTTGGAGGGCCTTGAGAGATGAATCAAGAGTCCCGCGGCG 121
QY 298 TGGAGAACTTCTGCAGAGACTTTTGAGCTGCAGAAAGTGTGTTACCTACCGCTCAACACCT 357
|||||
Db 122 TGGAGAACTTCTGCAGAGACTTTTGAGCTGCAGAAAGTGTGTTACCTACCGCTCAACACCT 181
QY 358 TCCTCTCTGGCGGCACTGCACCGGCTCATGCACACTACAAGCAGGTCTCTGGAGCGGCTGTGA 417
|||||
Db 182 TCCTCTCTGGCGGCACTGCACCGGCTCATGCACACTACAAGCAGGTCTCTGGAGCGGCTGTGA 241
QY 418 AACACCACCGCGGAGCCAGCGGACTTCAGGAGACTGCGAGCGGCTTTGGCAGAGATCA 477
|||||
Db 242 AACACCACCGCGGAGCCAGCGGACTTCAGGAGACTGCGGAGCGGCTNTGGCAGAGATCA 301
QY 478 CGGAGATGTTGGCAGAGCTCCACGGTACGATGATCAAGATGAGAGATTTCCAGAAGCTGC 537
|||||
Db 302 CGGAGATGTTGGCAGAGCTCCACGGTACGATGATCAAGATGAGAGATTTCCAGAAGCTGC 361
QY 538 ACGAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGTTCCGGGAAGGAGTTCA 597
|||||
Db 362 ACGAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGTTCCGGGAAGGAGTTCA 421
QY 598 TCCGTCGCGGAGCCCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 657
|||||
Db 422 TCCGTCGCGGAGCCCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 481
QY 658 TGTTCACAGAGCTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAG 717
|||||
Db 482 TGTTCACAGAGCTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAG 541
QY 718 TCCACGGGAGCTCCCGCTTATGGCATGACGATTGAGGAGAGCGAAGACGA 769
|||||
Db 542 TCCACGGGAGCTCCCGCTTATGGCATGACGATTGAGGAGAGCGAAGACGA 593

RESULT 7
Bg907778 890 bp mRNA linear EST 20-OCT-2000
LOCUS 601502061F1 NTH_MGC_70 Homo sapiens cDNA clone IMAGE:3903855 5',
DEFINITION mRNA sequence.
ACCESSION Bg907778
VERSION Bg907778.1 GI:10401682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 890)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9708 row: j column: 16
High quality sequence stop: 646.
Location/Qualifiers
```

FEATURES


```
Db 575 ACCAACCGCCGACGACCGCGACTTTCAGGGGACTGGCGAGCGCTTTGCGAGATCAGC 634
Qy 480 GAGATGGTGGCAGAGCTCCACCGTACGATGATCAAGATGAGAATTTCCAGAGCTGCAC 539
Db 635 GAGATGGTGGCAGAGCTCCACCGTACGATGATCAAGATGAGAATTTCCAGAGCTGCAC 594
Qy 540 GAACCTCAAGAAAGATTTGATGGCATGACATCTTGTGGTCCCGGGA---GGAGTT 595
Db 695 AACTCAAGAAAGATTTGATGGCATGACATCTTGTGGTCCCGGGAAGCGAGTTCC 754
Qy 596 CATCCGCTGGCGACGCTCAGCAA---GCTCTGGGGAAGGGCTCCAGCAGCGCATGT 651
Db 755 ATCCGCTGGCGACGCTTCAAGCAAGCTCCTCGGGGAAGGAGCTCCAGCAGCGCATG 814
Qy 652 TCTTCTGTTCAACGAGCTCCT-GCTATACAGAGCGCGGGGTGACGCGCTCCCAATCAG 710
Db 815 TCTTCTGTTCAAGAGCTCCTGGGTTACACAGCGCGGGGTGACAGCGTCCCAATCCA 874
Qy 711 TTTAAAGTCCAGCGGAGCTCC 732
Db 875 CGTTAAAGCAAGGCGAGTCCC 896

RESULT 9
BG750463 788 bp mRNA linear EST 15-MAY-2001
LOCUS 602709323F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4845892 5',
DEFINITION mRNA sequence.
ACCESSION BG750463
VERSION BG750463.1 GI:14061116
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 788)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI685 row: n column: 05
High quality sequence stop: 756.
Location/Qualifiers
1, 788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4845892"
/clone.lib="NIH_MGC_43"
/tissue.type="normal pigmented retinal epithelium"
/lab.host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"
BASE COUNT 196 a 210 c 229 g 153 t
ORIGIN

Query Match 72.9%; Score 560.6; DB 12; Length 788;
Best Local Similarity 94.7%; Pred. No. 2.le-137;
Matches 721; Conservative 0; Mismatches 24; Indels 16; Gaps 13;
```

```
Qy 24 GTGAGCAAGAGAGAGCCCATGCGGAAGCAGCTCAAAA-GTCTCATATATCCCGAATTTGA 82
Db 1 GTGAGCAAGAGAGAGAGCCCATGCGGAAGCAGCTGTCAACGCTCATATATCCCGAATTTGA 60
Qy 83 ACCTTTGGCAGAAA-TTTTCACTAATTTTCTCAAGGAAATTTGAGCAACAGCTTGCCTGT 141
Db 61 ACCTTTGGCAGAAAATATCATATATTTTCTCAAGGAAATTTGAGCAACAGCTTGCCTGT 120
Qy 142 GGAAGGCGCGCTCAA-ATGCCAAATCAGAGATTACCAAGAAATCGCG-GATGCTCATGCT 199
Db 121 GGAAGGCGCGTCAAGATGCCAAATCAGAGATTACCAAGAAATCGCGTGTATGCTATGCT 180
Qy 200 GAAGAACTTTCAGGGCATGAAGCACCTTGGCGCTACCTTGTGGAAGCAGCAGGAGCCCTT 259
Db 181 GAAGAACTTTCAGGGCATGAAGCACCTTGGCGCTACCTTGTGGAAGCAGCAGGAGCCCTT 240
Qy 260 GGAGGCCCTTGGAGATGGAAATCA-AGAGCTCCGGGGGCTGGAGAGACTTCTTCGAGAGACT 318
Db 241 GGAGGCCCTTGGAGATGGAAATCAGAGAGCTCCGGGGGCTGGAGAGACTTCTTCGAGAGACT 300
Qy 319 TTGAGCTGCAGAAAGTGTGTACCTA--CGGCTCAACACTTCTCCTCGGGCCACTGCA 376
Db 301 TTGAGCTGCAGAAAGTGTGTACCTAGCCGCTCAACACTTCTCCTCGGGCCACTGCA 360
Qy 377 CCGGCTCATGCACTA-CAAGCAGAGTCTTGGAGCGGC---TGTGCAAAACACACCCGCCGA 432
Db 361 CCGGCTCATGCACTATCAAGCAGAGTCTTGGAGCGGCTGTGGAAGACACCGACCCGCCGA 420
Qy 433 GCC-ACGCGCACTTTCAGGAGCTGCCGAGCGCTTTCGGCAGAGATCAGGAGATGGTGGA 491
Db 421 GCCGAGCGCACTTTCAGGAGCTGCCGAGCGCTTTCGGCAGAGATCAGGAGATGGTGGA 480
Qy 492 CAGCTCCAGGCTACGATGATCAAGATGGAGAAATTTCCAGAACTGCAGCAACTCAAGAAA 551
Db 481 CAGATCCAGGCTACGATGATCAAGATGGAGAAATTTCCAGAACTGCAGCAACTCAGAAA 540
Qy 552 GATTTGATTGGCATTTGACAATCTTGTGGTTCGGG-GAAGGGAGTTCATCCGCTGGGCGAG 610
Db 541 GATTCGATTGGCATTTGACAATCTTGTGGTTCGGGCAAGGGAGTTCATCCGCTGGGCGAG 600
Qy 611 CCTCAGCAAGCTCTCGGGGAA-GGGGCTCCAGCAG-CGGATGTTCTTCTGTTCAACAGC 658
Db 601 ACTCAGCAAGCTCTCGGGGAAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACAGC 660
Qy 669 GTCCTCTATACAGAGCGCGGGCTGACGGCTCCCAATCAGTTTAAAGTCCACGGGCGAG 728
Db 661 GTCCTCTATACAGAGCGCGGGGCTGACGG-CTCCAATCAGTTTAAAGTCCACGGGCGAG 719
Qy 729 CTCCTCTCTATGGCATGACGATTGAGGAGCGGAAGACGA 769
Db 720 CTCCTCTCTATGGCATGACGATTGAGGAGCGGAAGACGA 760

RESULT 10
BG331885/c 618 bp mRNA linear EST 17-MAY-2002
LOCUS BG331885/c
DEFINITION PM3-ET0268-100501-005-d11 ET0268 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG331885
VERSION BG331885.1 GI:20973153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 618)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
TITLE
```


Db 496 GAATCGCGGATGTCATGCTGAAGAACAATTCAGGGCATGAAGACACCTGGCGGCTCACCTGT 555
 QY 241 GGAAGCACAGCGAGCGCTTGGAGCCCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGG 300
 Db 556 GGAAGCACAGCGAGCGCTTGGAGCCCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGG 615
 QY 301 AGAATCTTCGAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCC 360
 Db 616 AGAATCTTCGAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCC 675
 QY 361 TCCTCGCGGCACTGCACCGGCTCATGCACCTACAGCAGGTCTCTGGAGCGGCTGTGCAAC 420
 Db 676 TCCTCGCGGCACTGCACCGGCTCATGCACCTACAGCAGGTCTCTGGAGCGGCTGTGCAAC 735
 QY 421 ACCACCGCGCG--AGCCAGCGGCTTACGGGACTCCGAGCGCG--TTTGGCAGAGATCA 477
 Db 736 ACCACCGCGCGGAGCGGCTTACGGGACTCCGAGCGCGCTTTTGGCAGAGATCA 795
 QY 478 C--GGAGATGGTGGACAGCTCCACGGTAGGATGATCAAGATGGAGAAATTTCCAGAGCT 535
 Db 796 CCGGAATGGTGGACAGCTCCACGGTAGGATGATCAAGATGGAGAAATTTCCCGAAGCT 855
 QY 536 GCACGAACTCAAGAAAGATTGATTGGCATTGACAATCTTGTGG 579
 Db 856 GGCCCAACTCAGAAAGATTGATTGGCATTGACAATCTTGTGG 899

RESULT 12
 BM705217
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 9704477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse
 Location/Qualifiers
 1. .496
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-Cil-afg-c-04-0-UI"
 /clone_lib="UI-E-Cil"
 /tissue_type="RPE and Choroid"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phase resistant)"
 /note="Organ: eye; Vector: p7T3-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-Cil is a normalized cDNA library containing the
 following tissue(s): RPE and Choroid. The library was
 constructed according to Bonaldo, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into p7T3-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is ACCTA.
 This library was created for the program, Gene Discovery
 in the Visual System, supported by National Eye Institute
 (NEI)."

BASE COUNT 133 a 131 c 133 g 99 t
 ORIGIN

Query Match 64.3%; Score 494.4; DB 14; Length 496;
 Best Local Similarity 99.8%; Pred. No. 5.3e-120;
 Matches 495; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 74 GAATTTTGAACCTTTGCACAAATTTTCATATAATTTTCTCAAGGAAATTTGAGCAAGCACT 133
 Db 1 GAATTTTGAACCTTTGCACAAATTTTCATATAATTTTCTCAAGGAAATTTGAGCAAGCACT 60

QY 134 TGCCTCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCCGCGATGT 193
 Db 61 TGCCTCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCCGCGATGT 120

QY 194 CATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGA 253
 Db 121 CATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCTGTGGAAGTACAGCGA 180

QY 254 GGCTTTGGAGGCGCTCGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAG 313
 Db 181 GGCTTTGGAGGCGCTCGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAG 240

QY 314 AGACTTTGAGCTCGACAGAGGTGTACCTACCGCTCAACACCTTCTCTCGGCGCACT 373
 Db 241 AGACTTTGAGCTCGACAGAGGTGTGTACCTACCGCTCAACACCTTCTCTCGGCGCACT 300

QY 374 GCACCGGCTCATGCTACATCAAGCAGGTCTGTGGAGCGGCTGTGCAACACACCCGCCGAG 433
 Db 301 GCACCGGCTCATGCTACATCAAGCAGGTCTGTGGAGCGGCTGTGCAACACACCCGCCGAG 360

QY 434 CCACGCGGCTTTCAGGAGCTGCCGAGCGCTTTGGCAGAGATCACGAGATGTGTGGCACA 493
 Db 361 CCACGCGGCTTTCAGGAGCTGCCGAGCGCTTTGGCAGAGATCACGAGATGTGTGGCACA 420

QY 494 GCTCCACGGTACGATGATCAAGATGGAGATTTCCAGAGCTGCACGAACTCAAGAAGA 553
 Db 421 GCTCCACGGTACGATGATCAAGATGGAGATTTCCAGAGCTGCACGAACTCAAGAAGA 480

QY 554 TTTGATTGGCATTGAC 569
 Db 481 TTTGATTGGCATTGAC 496

RESULT 13
 AL121548
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 489)
 Ottenwaelder,B., Oberwater,B., Mewes,H.W., Gassenhuber,J. and
 Wiemann,S.
 EST (Ottenwaelder, et al.)
 JOURNAL
 COMMENT
 Contact: Ottenwaelder B

AL121548 489 bp mRNA linear EST 25-FEB-2000
 DKFp762L158_r1 762 (synonym: hm12) Homo sapiens cDNA clone
 DKFp762L158 5', mRNA sequence.
 AL121548
 AL121548.1 GI:5927549
 EST.
 human.

MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Medigenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. No s1 sequence
available.
This clone (DKFZp762L158) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
 source
 1..489
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp762L158"
 /clone_lib="762 (synonym: hmel2)"
 /tissue_type="melanoma (MeWo cell line)"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 121 a 134 c 143 g 91 t
ORIGIN
Query Match 61.8%; Score 475.4; DB 9; Length 489;
Best Local Similarity 99.6%; Pred. No. 5.5e-115;
Matches 487; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 125 GCACGACTTGGCTTGGAGGCGCTCAATGCCCCAAATCAGAGATTACCAAGAAT 184
 |||||
Db 1 GCAACGACTTGGCTTGGAGGCGCTCAATGCCCCAAATCAGAGATTACCAAGAAT 60
QY 185 CGCGATGTCATGCTGAAGCAATTCAGGCGCATGAAGCACTGGCGCTCCTGTGGAA 244
 |||||
Db 61 CGCGATGTCATGCTGAAGCAATTCAGGCGCATGAAGCACTGGCGCTCCTGTGGAA 120
QY 245 GCACGCGAGGCTTGGAGGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGCTGGAGAA 304
 |||||
Db 121 GCACGCGAGGCTTGGAGGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGCTGGAGAA 180
QY 305 CTTCTGCAGAGCTTGAAGCTGAGAGAGTGTGTACCTACCGCTCAACACCTTCTCTCT 364
 |||||
Db 181 CTTCTGCAGAGCTTGAAGCTGAGAGAGTGTGTACCTACCGCTCAACACCTTCTCTCT 240
QY 365 GGGGCACTGCACCGGCTCATGCACATACAGCAGGTCCTGGAGCGGCTGTGCAACACCA 424
 |||||
Db 241 GCGGCCACTGCACCGGCTCATGCATACAGCAGGTCCTGGAGCGGCTGTGCAACACCA 300
QY 425 CCGCGGAGCCAGCGGACTTCAGGAGCTGCGGAGCGGCTTTGGCAGAGATCACGAGAT 484
 |||||
Db 301 CCGCGGAGCCAGCGGACTTCAGGAGCTGCGGAGCGGCTTTGGCAGAGATCACGAGAT 360
QY 485 GGTGGCAGAGCTCCAGGTTACGATGATCAAGATGAGAAATTTCCAGAGCTGCAGAACT 544
 |||||
Db 361 GGTGGCAGAGCTCCAGGTTACGATGATCAAGATGAGAAATTTCCAGAGCTGCAGAACT 420
QY 545 CAAGAAGATTTGATTGGCATTGACAACTCTGTGGTCCGGGAAGGAGTTCA-TCCGTC 603
 |||||
Db 421 CAAGAAGATTTGATTGGCATTGACAACTCTGTGGTCCGGGAAGGAGTTCA-TCCGTC 480
QY 604 TGGCGAGCC 612
 |||||
Db 481 TGGCGAGCC 489
RESULT 14
BG420356
LOCUS
DEFINITION 602448377F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4586971 5',
 861 bp mRNA linear EST 14-MAR-2001
 BG420356
ACCESSION BG420356
VERSION BG420356.1 GI:13326862
KEYWORDS EST, .

SOURCE
ORGANISM human.
 Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DPF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1316 row: i column: 20
 High quality sequence stop: 569.
FEATURES
 source
 1..861
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4586971"
 /clone_lib="NIH_MGC_14"
 /tissue_type="renal cell adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 188 a 249 c 268 g 156 t
ORIGIN
Query Match 60.2%; Score 463; DB 12; Length 861;
Best Local Similarity 100.0%; Pred. No. 1.4e-111; Mismatches 0; Gaps 0;
Matches 463; Conservative 0; Indels 0;
QY 307 TCTGCAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCTCTCTGC 366
 |||||
Db 3 TCTGCAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCTCTCTGC 62
QY 367 GGCCTACTGACCGGCTCATGCTACAGCAGGTCTGGAGCGGCTGTGCAACACCAACC 426
 |||||
Db 63 GGCCTACTGACCGGCTCATGCTACAGCAGGTCTGGAGCGGCTGTGCAACACCAACC 122
QY 427 CGCCGAGCCAGCCGCTTTCAGGAGCTGCGGAGCGCTTTGGCAGAGATCACGGAGATGG 486
 |||||
Db 123 CGCCGAGCCAGCCGCTTTCAGGAGCTGCGGAGCGCTTTGGCAGAGATCACGGAGATGG 182
QY 487 TGGCAGAGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAGCTGCAGAACTCA 546
 |||||
Db 183 TGGCAGAGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAGCTGCAGAACTCA 242
QY 547 AGAAGATTTGATTGGCATTGACAACTCTGTGGTCCCGGAAGGAGTTTCATCCGTCGG 606
 |||||
Db 243 AGAAGATTTGATTGGCATTGACAACTCTGTGGTCCCGGAAGGAGTTTCATCCGTCGG 302
QY 607 GCAGCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTCTTCTCTGTTCAAGC 666
 |||||
Db 303 GCAGCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTCTTCTCTGTTCAAGC 362
QY 667 AGCTCCTGCTATACACGAGCCGGGGCTCACCGCTTCCCAATCAGTTTAAAGTCCACGGGC 726
 |||||
Db 363 AGCTCCTGCTATACACGAGCCGGGGCTCACCGCTTCCCAATCAGTTTAAAGTCCACGGGC 422
QY 727 AGCTCCCGCTCTATGGCATGACGATTGAGGAGGAGCGAAGACGA 769
 |||||
Db 423 AGCTCCCGCTCTATGGCATGACGATTGAGGAGGAGCGAAGACGA 465

RESULT 15

BM551766
LOCUS BM551766 1026 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6575294 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5479350
5', mRNA sequence.
ACCESSION BM551766
VERSION BM551766.1 GI:18789071
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1026)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

JOURNAL

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LCM2000 row: h column: 07

High quality sequence stop: 675.

FEATURES

source

Location/Qualifiers

1..1026

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5479350"

/clone_lib="NIH_MGC_98"

/tissue_type="astrocytoma grade IV, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GCGACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

BASE COUNT

227 a 309 c 296 g 194 t

ORIGIN

Query Match 59.4%; Score 457; DB 13; Length 1026;
Best Local Similarity 100.0%; Pred. No. 5.8e-110;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 GAGACTTTGAGCTGAGAGGTGTTACCTACCGCTCAACACCTTCCTCTGGCGGCAC 372
|||||

Db 1 GAGACTTTGAGCTGAGAGGTGTTACCTACCGCTCAACACCTTCCTCTGGCGGCAC 60
|||||

QY 373 TGCACCGGCTCATGCACATACAGAGGTGCTGGAGCGGTGTGCAACACCAACCCGCCGA 432
|||||

Db 61 TGCACCGGCTCATGCACATACAGAGGTGCTGGAGCGGTGTGCAACACCAACCCGCCGA 120
|||||

QY 433 GCCACCGGCTCATGCACATACAGAGGTGCTGGAGCGGTGTGCAACACCAACCCGCCGA 492
|||||

Db 121 GCCACCGGCTCATGCACATACAGAGGTGCTGGAGCGGTGTGCAACACCAACCCGCCGA 180
|||||

QY 493 AGCTCCAGGTCAGATGATCAAGATGGAGATTTCAGAGCTGCACGAACTCAAGAAG 552
|||||

Db 181 AGCTCCAGGTCAGATGATCAAGATGGAGATTTCAGAGCTGCACGAACTCAAGAAG 240
|||||

QY 553 ATTTGATTGGCAATTCACAATCTTGTGGTCCGGGAAGGAGTTTCATCGCTGTGGCAGCC 612
|||||

Db 241 ATTTGATTGGCAATTCACAATCTTGTGGTCCGGGAAGGAGTTTCATCGCTGTGGCAGCC 300
|||||

QY 613 TCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTCC 672
|||||

Db 301 TCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTCC 360
|||||

QY 673 TGCTATACACGAGCGGGGGCTGACGGCCTCCAATCAGTTTAAAGTCCACGGCGAGCTCC 732
|||||

Db 361 TGCTATACACGAGCGGGGGCTGACGGCCTCCAATCAGTTTAAAGTCCACGGCGAGCTCC 420
|||||

QY 733 CGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 769
|||||

Db 421 CGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 457
|||||

Search completed: December 12, 2002, 06:39:28
Job time : 2443 secs

THIS PAGE BLANK (USPTO)